DOCUMENT RESUME

ED 380 492 TM 022 853

AUTHOR	Ekstrom, Ruth B.; Villegas, Ana Maria
TITLE	College Grades: An Exploratory Study of Policies and Practices. College Board Report No. 94-1.
INSTITUTION	College Board, New York, NY.; Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RR-94-23
PUB DATE	94
NOTE	39p.
AVAILABLE FROM	College Board Publications, Box 886, New York, NY 10101-0886 ($15).
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Change; *College Faculty; Criterion Referenced Tests; Educational Policy; Educational Practices; *Educational Trends; *Grades (Scholastic); *Grading; Higher Education; Informal Assessment; Norm Referenced Tests; *Standards; Student Evaluation; Student Evaluation of Teacher Performance; Teacher Expectations of Students; Trend Analysis

ABSTRACT

Policies and practices related to grading at 14 colleges and universities (8 public and 6 private) and how they have changed between 1980 and 1990 were studied along with the grading orientations and practices of over 500 faculty members. Changes between 1980 and 1990 that may have affected grades include greater prescription of the curriculum, greater differentiation in grading systems, and increased use of student evaluations of faculty members. While none of the department chairpersons said there were specific departmental grading policies, about one quarter of the faculty felt that there was a policy of grading against specific standards, and about two-thirds said that their departments expected them to grade against specific standards, even though there was no defined policy. Some faculty members perceived grades as formal and objective, while others thought they could not be reduced to objective measures. Most believed that grades are higher than they used to be because faculty expects less of students today. Eighty-one percent of faculty said they sometimes used a criterion-referenced approach, but only 64% said they used it most often. Twenty-nine percent used a norm-referenced approach most often. Twenty-four tables present study findings. (Contains 91 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

College Board Report No. 94-1

College Grades: An Exploratory Study of Policies and Practices

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it
☐ Minor changes have been made to improve reproduction quality
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
L.E. Gladieux
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

ED 380 492

TM 022 853

RUTH B. EKSTROM and ANA MARÍA VILLEGAS

The College Board
Educational Excellence for *All* Students

College Board Report No. 94-1
ETS RR No. 94-23

College Grades: An Exploratory Study of Policies and Practices

RUTH B. EKSTROM and ANA MARÍA VILLEGAS
with the assistance of BARBARA A. BRUSCHI
and FREDERICK A. CLINE

College Entrance Examination Board, New York, 1994

Acknowledgments

The authors wish to thank the colleges and universities, department chairpersons, and faculty members who provided the information used in this report. We also wish to thank Gwen Boodoo, Len Ramist, and Ken Wilson for their helpful comments on an earlier draft.

Ruth Ekstrom is acting executive director and a principal research scientist in the Education Policy Research Division at ETS.
Ana Maria Villegas is a research scientist at ETS.
Barbara A. Bruschi is a senior research assistant at ETS.
Frederick A. Cline is a research assistant at ETS.

Researchers are encouraged to freely express their professional judgment. Therefore, points of view or opinions stated in College Board Reports do not necessarily represent official College Board position or policy.

The College Board is a national nonprofit association that champions educational excellence for all students through the ongoing collaboration of more than 2,900 member schools, colleges, universities, education systems, and organizations. The Board promotes—by means of responsive forums, research, programs, and policy development—universal access to high standards of learning, equity of opportunity, and sufficient financial support so that every student is prepared for success in college and work.

Additional copies of this report may be obtained from College Board Publications, Box 886, New York, New York 10101-0886. The price is $15.

Copyright © 1994 by College Entrance Examination Board. All rights reserved. College Board, SAT, and the acorn logo are registered trademarks of the College Entrance Examination Board.

Printed in the United States of America.

Contents

Abstract 1

Introduction 2
- Background and Rationale 2
- Institutional Policies and Practices 3
- Variations in Grading by Department or Discipline 5
- Faculty Grading Practices 6
- Changes in Student Achievement Level and in the Perceived Importance of Grades 7

Method 8
- Instruments 8
- Institutional Sample 9
- Department and Faculty Sample 10
- Data Collection 10

Results 11
- Institutional Analysis 11
- Department Chairperson Analysis 13
- Individual Faculty Member Analysis 14
- Departmental Information 14
- Faculty Attitudes about Grading 16

Discussion and Conclusions 27
- Changes Between 1980 and 1990 27
- Institutional and Departmental Influences on Grades 28
- Overview of Faculty Grading Philosophies and Practices 29
- Departmental Variations in Grading 29
- Additional Analyses and Future Research 30
- Grading in the Mid-1990s 30

References 31

Tables

1. Institutional Selectivity in 1990: Self-Reported Versus *College Handbook* 10
2. Number of Institutions Using Various Grading Systems in 1990 and in 1980 11
3. Summary of Institutional Grading Policies in 1990 and in 1980 12
4. Institutional Uses of Student Evaluations of Faculty in 1990 and in 1980 13
5. Percentage of Department Chairs Reviewing Faculty Grades 13
6. Percentage of Department Chairs Reporting Formal and Informal Discussions of Grading, by Department and Institutional Characteristics 14
7. Number of Faculty Responding by Field and Institutional Type 14
8. Percentage of Responding Faculty with Various Background Characteristics 15
9. Percentage of Responding Faculty Reporting Department Efforts to Raise Standards, by Field 16
10. Percentage of Responding Faculty Reporting Various Departmental Policies and Expected Practices Related to Grading 16
11. Faculty Attitudes about Grading, Means by Field and for Total Group, and Probability of *F*-Test Ratio for Significant Differences Across Fields 17
12. Percentage of Faculty Who Had Changed Their Grading Philosophy since They Began Teaching 18
13. Faculty Opinions about Purposes of Grading, Means by Field and for Total Group, and Probability of *F*-Test Ratio for Significant Differences Across Fields 18
14. Percentage of Responding Faculty with Various Opinions about the Most Important Factors in Grade Inflation 19
15. Grading Practices of Responding Faculty and Probability of *F*-Test Ratio for Significant Differences Across Fields 20
16. Grading System Most Often Used by Responding Faculty and Probability of *F*-Test Ratio for Significant Differences Across Departments 20

17. Faculty Perceptions of How Grades in Their Introductory Courses Compared with Those of Other Faculty 21

18. Importance Given to Various Factors in Assigning Grades in Introductory and Advanced Courses 21

19. Departmental Means for Importance Given to Various Factors When Assigning Grades in Introductory Courses and Probability of *F*-Test Ratio for Significant Differences Across Fields 22

20. Factors with Varying Levels of Importance in Assigning Grades in Introductory Undergraduate Courses 23

21. Departmental Means for Importance Given to Various Factors When Assigning Grades in Advanced Courses and Probability of *F*-Test Ratio for Significant Differences Across Fields 24

22. Factors with Varying Levels of Importance in Assigning Grades in Advanced Undergraduate Courses 25

23. Importance Given to Various Factors in Assigning Grades in Laboratory, Fieldwork, and Internship Courses Versus Other Courses 26

24. Types of Assessment Used in Introductory Undergraduate Courses and Probability of *F*-Test Ratio for Significant Differences Across Fields 27

Abstract

This study examined policies and practices related to grading at 14 colleges and universities and how they had changed between 1980 and 1990. It also examined the grading orientation and practices of over 500 faculty members in the business, chemistry, education, English, history, mathematics, and psychology departments of these institutions.

The study sought information to answer four questions:

1. What are the current institutional and/or departmental grading policies and practices?
2. Were there changes in these policies and practices between 1980 and 1990 and, if so, what was the nature of the changes?
3. What are the current grading orientation and practices of faculty who teach undergraduate courses? Have these changed over time?
4. Do faculty grading orientation and practices differ across departments and, if so, how?

Information was collected from 14 colleges and universities, 8 public and 6 privately controlled. Within each institution, information was collected from department chairs and faculty in each of the designated departments that was established at that institution. Usable questionnaire responses were received from 58 department chairs (68 percent response rate) and from 542 faculty members (25 percent response rate).

Institutional changes between 1980 and 1990 that may have affected grades included greater prescription of the curriculum, greater differentiation in grading systems, and increased use of student evaluations of faculty members.

While none of the department chairpersons said there were specific departmental grading policies, faculty had a very different perspective. Approximately a quarter of the faculty said their department had a policy of grading against specific standards, 14 percent of these said there was a policy that all students be given an A or B in honors courses, and 11 percent said there was a policy that attendance should be a factor in course grades. Additionally, about two-thirds of the faculty said that, although not specific policy, their department expected them to grade students against specific standards. About half of these faculty said their department expected them to grade students relative to the overall performance of the class.

Only about a third of the department chairpersons reported having formal meetings to discuss grading, but informal meetings about grading were reported by 75 percent. Fifty-six percent of the responding faculty said their department had tried to raise standards during the decade between 1980 and 1990.

Faculty orientations toward grading could be categorized into two approaches, one viewing grades as formal and objective, the other insisting that grades cannot be reduced to a set of objective measures. There were significant differences across departments, with chemistry, mathematics, and psychology faculty more likely to subscribe to the "objective" view while English, education, and history faculty tended to favor the "nonobjective" view. Faculty in business appeared to have diverse attitudes toward grading and could not be easily categorized into either group.

Forty-three percent of the responding faculty said their grading philosophy had changed since they began to teach. Faculty believed that the meaning of a grade varies more across disciplines than across institutions and that the major reason why grades today are higher than they were 20 years ago is because faculty now expect less of students.

When asked about the purpose of grades, responding faculty saw the primary purposes as providing feedback to students, providing information about students to graduate or professional schools, motivating students to do good work, and helping the college or university make decisions about students.

Faculty were asked if they ever used each of three grading approaches. Eighty-one percent said they sometimes used a criterion-referenced approach, 57 percent a norm-referenced approach, and 44 percent a self-referenced approach. When asked which one they used most often, 64 percent named the criterion-referenced approach, 29 percent the norm-referenced approach, and 8 percent the self-referenced approach.

When asked about the importance they gave to various components of grades when assigning grades in introductory and advanced courses, tests and quizzes were rated as most important for introductory courses, papers and written assignments for advanced courses. Other factors that took on increased importance in grading advanced courses included oral reports, creativity, class participation, group projects, and subject-specific skills and techniques. There were many significant differences across departments in the importance assigned to different components of grades.

Responding faculty were more likely to use essay than multiple-choice tests in their introductory courses, but again, there were significant differences by department.

Introduction

It is common knowledge among members of the higher education community that grading practices lack uniformity; grade distributions vary considerably among sections of the same course and across instructors, departments, and institutions. When Lewis, Dexter, and Smith (1978) studied the grading procedures of college English teachers, they found 10 different patterns of information used to arrive at grades. Small wonder then that those analyzing grades have made comments such as "The unreliability of grades within departments and the variability of grading standards across departments is apparent to all who have looked into these matters" (Fishman 1958, 341).

This report explores some of the factors involved in college grades. It describes a study of current undergraduate course grading policies and practices in 14 colleges and universities. It also provides some insights into how these grading policies and practices may have changed between 1980 and 1990.

Grading is clearly a complex issue. Grades may be affected by institutional and departmental policies as well as by the types of individuals doing the grading, the grading orientation and practices of these individuals, and the kinds of students who enroll in a given institution or a given course. Most of the existing research on grading has looked at only one or two of these factors. To get a more comprehensive picture, information was collected at the institutional, departmental, and individual faculty levels to determine how they interact and which appear to be primary influences.

Background and Rationale

Grades and grading practices are topics of widespread interest, as well as some controversy, in higher education. Underlying the discussion of grades has been a concern for standards, which at times have been considered "too stringent" and at other times "too lax." For example, during the 1950s and 1960s there was concern over the increasing rigor of grading standards (see Webb 1959; Aiken 1963; Hills 1964; Hills and Klock 1965; Miller 1965; Hills 1967; Hills and Gladney 1968; Wilson 1970). During the past two decades, especially in the 1970s, the concern was the declining value of grades, or what has come to be known as "grade inflation" (see Juola 1974; Etzioni 1975; Suslow 1977; Bejar and Blew 1981; Milton, Pollio, and Eison 1986; Summerville, Ridley, and Maris 1988). In addition to the possibility that grade inflation may represent an erosion of academic standards or be a counteraction to the more stringent standards used in the 1950s and 1960s, it may also represent real increases in student achievement, demographic changes, or institutional grading policies that permit grade-point averages (GPAs) to rise even if grading standards remain unchanged (Birnbaum 1977). It is likely that, on average, grade inflation was a phenomenon of the 1970s; pairwise matches of validity studies (Ramist and Weiss 1990) revealed a huge grade inflation in the 1970s but none thereafter.

Sabot and Wakeman-Linn (1991) reported evidence from nine colleges and universities that grade inflation during the past 25 years has resulted in an increasing disparity of grades and grading policies within institutions, splitting many into high- and low-grading departments. The authors argue that these differences affect course choice and make it difficult to attract students to fields like mathematics and science in which grades tend to be lower than in the arts and humanities. Students use grades to determine if they should take further courses in a field and to make decisions about the choice of a major and/or a career (McKeachie 1986).

Another aspect of the current concern about college grades is the question of their adequacy as an indicator of college achievement. Since the chief function of grades is to convey information about students, unexplained changes and variations in grading policies raise questions about their validity. As Geisinger (1982) pointed out in his review of marking systems, "numerous studies have demonstrated that individual teachers differ in the kinds of marks they assign and that these differences are stable." Duke (1983) found differences in grading practices among the colleges of one university, among the various disciplines, and among the instructors. These differences have led some to question the almost universal use of grades in making educational admission and employment decisions. (See Hoyt 1966, 1970; Humphreys 1968; O'Leary, 1980.)

A related concern about grades centers on their relationship to test scores and their predictability, especially as used in college and graduate school admission. Various grade adjustment methods have been employed; for a summary of these see Young (1993). Lewis, Dexter, and Smith (1978) found that grades from English teachers who relied on each of several different sets of factors (marking patterns) in assigning grades correlated differently with test scores. In considering factors that might influence the relationship between test scores and freshman grades, Willingham (1990a) listed the following variations that might be observed in the criterion (grades):

- characteristics of grades (courses included, grading practices);

- meaning of grades (skills or type of performance evaluated);
- grading variations (due to different standards or performances); and
- educational practices (remediation, placement).

Willingham suggested a variety of factors that may have altered the predictability of the freshman GPA including: rigor of the grading process, grade inflation, variation in grading standards across courses, the incidence of remedial and advanced course work, diversity in the curriculum, changes in what is rewarded in grading, changes in the competencies required of students, and changes in program enrollment.

The predictability of college grades is also related to institutional characteristics (Baird 1983). Although some have argued that grades are less predictable in institutions that enroll students with a wide range of academic ability and in institutions that have a very diverse curriculum, correlations between freshman grade-point average (FGPA) and standard predictors tend to be higher at institutions that enroll students with a wide range of academic ability because there is less restriction of range. After correcting for such restriction there is some evidence of very slightly reduced correlations at such institutions (Ramist and Weiss 1990). Sabot and Wakeman-Linn (1991) found considerable differences across departments in the extent to which grades in introductory courses could be predicted by other indicators, including ability measures such as the SAT. They found much better prediction of grades in the low-grading departments than in the high-grading departments. There have also been concerns about bias associated with gender and race in the prediction of college performance. (See, for example, Hogrebe et al. 1983; Elliott and Strenta 1988; McCornack and McLeod 1988.)

The discussion now turns to a consideration of how college grades may be affected by: (1) institutional policies and practices; (2) variations in grading by department or discipline; (3) faculty attitudes about grading and faculty grading practices; and (4) changes in student achievement level and in the perceived importance of grades.

Institutional Policies and Practices

Students' GPAs may be subject to influence by a variety of institutional policies and practices, including admission policies, required courses, how course grades are recorded, which courses and grades are included in the GPA, the extent to which part-time faculty are used, and the way that student course evaluations are used. It is possible that institutional changes in these areas over the past two decades have affected the GPA. A survey of changes in grading practices between 1973–74 and 1978–79 (Collins and Nickel 1979) found that policies were more stringent in 1978–79 concerning the use of nontraditional grading practices, makeup of incompletes, withdrawal policies, and policies related to remaining in good standing, being placed on probation, or being suspended. But there were also more institutions reporting unlimited opportunities to repeat courses, not recording grades from a "bad" semester, and using the highest grade earned in a course to calculate the GPA. Unfortunately this area has received relatively little recent research attention.

Changing admission standards at some colleges and universities during the past two decades may have affected GPAs. Students who were admitted under lower standards may have been more likely to avoid courses known to have strict grading standards (Goldman et al. 1974). Supporting this argument, Prather, Smith, and Kodras (1979) stated that rising cumulative GPAs are not caused by a lowering of grading standards in individual courses; rather, they are the result of changes in student course enrollment patterns over time. Specifically, students are moving away from traditional courses, especially those known to be stiffly graded, to newer degree programs with more lenient grading standards. Hills (1964) found that when one liberal arts college raised its admission standards, students of equal academic ability (as indicated by their admission credentials) then tended to receive lower grades. He hypothesized that this was due to a faculty commitment to assign grades to a class on a specified percentage basis, i.e., 15 percent A's, 40 percent B's, etc.

Some attention has been given to the impact of curricular changes on college grades. Studies comparing general education requirements across time provide some insights into trends. In the 1960s, general education requirements represented 43 percent of the total degree requirements (Dressel and DiLisle 1969). By the early 1970s this had dropped to 33 percent (Blackburn et al. 1976), but it rose to 38 percent by the late 1980s (Toombs and Fairweather 1989). In their extensive national survey of undergraduate institutions, Blackburn and his colleagues (1976) identified a radical change away from the traditional core curriculum toward more specialized and narrow fields of study. In more recent years, there appears to have been an increase in general education requirements, most notably in mathematics, but also in other areas. Approximately 30 percent of colleges and universities had a mathematics general education requirement in the 1960s. This declined to 20

percent in the 1970s. However, Lewis and Farris (1989) found that 46 percent of four-year colleges and universities had such a requirement in 1983 and 59 percent in 1988. Toombs and Fairweather (1989) reported that 65 percent of the institutions they surveyed had such a requirement in 1988. Eliminating core curriculum requirements, as occurred during the 1970s, has been linked to inflated cumulative GPAs (Suddarth 1975; Potter 1979). However, the effect of the more recent changes in general education requirements is unknown. The Lewis and Farris study found that general education requirements represented a larger proportion of total degree requirements at comprehensive and liberal arts institutions than at research and doctoral institutions (using the Carnegie classifications).

Nontraditional courses compound the impact of curriculum change. Carney, Isakson, and Ellsworth (1978) found that the increased use of field experiences, internships, and other nontraditional course formats associated with new areas of specialization contributed even further to inflated cumulative GPAs because students generally received higher grades in such courses. Stein and Guthrie (1990) also found that students received higher grades in internship and field experience courses than in regular courses and those emphasizing laboratory work.

The type(s) of grade reporting system used by an institution appears to affect GPA. There is evidence that students in pass/fail courses would have earned lower grades had they been graded on the traditional basis (Felder 1979; Geisinger and Rabinowitz 1979; Potter 1979). Increasing the degree of differentiation in a grading system, such as by adding pluses and minuses to letter grades, may inflate the GPA. For example, Potter (1979) found that adding pluses and minuses to traditional letter grades resulted in higher grades than when the traditional five-letter system was used alone.

Institutional policies regarding how the GPA is computed will, of course, have an impact on grading outcomes. Research in this area has focused primarily on policies related to student withdrawal from courses and use of incompletes. Potter (1979) found that liberalized incomplete and withdrawal policies at one institution resulted in dramatically higher cumulative GPAs. Geisinger and Rabinowitz (1979) reported that students tended to withdraw from courses when their grades were low in order to protect their GPAs. A longitudinal study of policies concerning making up incompletes and withdrawals from courses found that the policies were slightly more strict in 1978 than they had been five years earlier (Collins and Nickel 1979); this trend, the researchers concluded, would help reduce grade inflation.

Duke concluded that, among the top students graduating from one institution in 1978, several had obtained their high standing because of administrative rules related to transfer credits, readmission, pass/fail courses, and repeated courses, which had a confounding impact on their GPAs. "The net result was that the top 19 students attained their high GPAs by decidedly nonequivalent accounting procedures. Hence, the resulting rankings were limited in reliability and validity and led to inequitable comparisons" (1983, 1049).

Institutions and departments vary in their policies and practices concerning the use of part-time faculty. Of late the use of part-time faculty has been increasing. This may be a response to the rapid growth in some disciplines. It is also often a means of dealing with tight budgets. One recent national estimate is that 35 to 40 percent of faculty hold part-time appointments (Mangan 1991). Prather (1976) found that part-time faculty tended to give higher grades than full-time faculty. One hypothesis is that part-time faculty are not sufficiently integrated into departmental activities to learn the grading norms used by their full-time colleagues. Another possibility is that part-time faculty, especially in applied fields such as business, may emphasize competencies valued more highly in the workplace than in academe.

Institutional and departmental use of students' course evaluations, especially when they are considered in faculty retention, tenure, and promotion decisions, may also be exerting upward pressure on grades. Faculty who give higher grades tend to receive more favorable evaluations from their students (Carney et al. 1978; Longstreth 1979; Johnson and Beck 1988; Rasbow and Hernandez 1988; Zangenehzadeh 1988; Nimmer and Stone 1991). This suggests that, if faculty are aware that student evaluations will be used to make decisions about retention, promotion, tenure, and salary, they may be afraid to grade too harshly lest they receive unfavorable reviews and jeopardize their careers. We would anticipate that faculty concern about student evaluations would increase in periods when college faculty have difficulty finding and keeping jobs. The value of student ratings as well as their potential for misuse has been discussed by Aleamoni (1987).

Still another institutional factor that may have affected grades in recent years is an administrative effort to deal with perceived or actual grade inflation. Anecdotal evidence suggests that some college administrators may be reviewing average grades in courses and calling faculty to task if their grades are noticeably higher than average. Other administrators are reported to be imposing practices such as "grading on the curve" as a

means of combating grade inflation or to create the appearance of high institutional standards.

Variations in Grading by Department or Discipline

One of the most consistent findings in the literature on grading is the variation from department to department or discipline to discipline (see Juola 1968; Goldman, Schmidt, Hewitt, and Fischer 1974; Goldman and Slaughter 1976; Prather, Smith, and Kodras 1979; Duke 1983; Willingham 1985; Strenta, and Elliott 1987; Summerville, Ridley, and Maris 1988). The same disciplines tend to be characterized by high or low grades when comparisons are made across institutions (Elliott and Strenta 1988; Summerville, Ridley, and Maris 1988; Sabot and Wakeman-Linn 1991). There is less agreement about whether or not these differences have changed over time. Duke (1983) reported that they appeared to be stable across time but Sabot and Wakeman-Linn (1991) found increasing divergence. College grades tend to be lower in the physical and biological sciences and higher in education and in the arts and humanities (Duke 1983; Boli, Katchadourian, and Mahoney 1984; Willingham 1990b; Sabot and Wakeman-Linn 1991).

This has become a matter for concern since students with comparable abilities may end up with widely differing GPAs depending on the mix of courses they take. Several studies have suggested adjusting, or changing the use of GPAs to take these departmental differences into consideration. For example, Duke (1983) proposed use of centile grades to provide a measure of relative achievement. Elliott and Strenta (1988) have shown that the use of an index of differential grading when predicting GPA not only improves overall GPA prediction but also improves prediction for minority students and reduces apparent underprediction for women. However, McCornack and McLeod (1988) reported that in most of the large introductory courses they studied, no gender bias was found that would hold up through cross-validation in another semester.

One hypothesis to explain departmental differences in grading patterns is the "two cultures" phenomenon, whereby different standards apply in the sciences and other fields that emphasize knowledge of specific factors and cumulative course content than apply in the humanities. These differences may also be related to the types of assessment used (Oh 1976; Kodras and Prather 1978; Chase and Wakefield 1984). Faculty in more data-oriented disciplines, such as the physical sciences, tend to emphasize objective tests and quizzes and are less willing to adjust their grading standards to accommodate individual student characteristics. In contrast, faculty in the humanities, fine arts, and education are less likely to use objective tests and more likely to consider factors such as effort and attitude when they assign grades.

Etzioni (1975) has theorized that views of knowledge, held both within disciplines and by individual faculty members, profoundly influence grading practices. According to Etzioni, faculty members who see knowledge as a formal and objective system tend to have higher grading standards than faculty who view knowledge as socially constructed and relative. Using this argument, Etzioni attributed increases in GPA during the 1970s to a broader acceptance of the socially constructed and relative view of knowledge on the part of college faculty. Etzioni's (1961) organizational theories have been used by Bromley, Busching, Oliver, and Szozda to compare grading in normative departments (those in which faculty are concerned with reaching out to students and transforming them) with grading in utilitarian departments (those that emphasize presentation of the subject matter). They found significant differences in grading practices and argued that attempts to standardize grades across disciplines "may force uniformity which thwarts legitimate diversity in the academic goals of various departments" (1981, 434).

A closely related finding is the significant difference in GPAs from school to school within larger institutions (Duke 1983; Stein and Guthrie 1990). This may be related to differences in admission policies, academic standards, departmental policies, and/or required courses across schools.

Class size and course level may also be related to grading differences, both across and within departments. Grades appear to be inversely related to class size (Boli, Katchadourian, and Mahoney 1984; Dickson 1984; Stein and Guthrie 1990), with the effect strongest in the freshman year, according to Dickson. Stein and Guthrie argued that smaller classes permit higher quality instruction, including better evaluation practices. Geisinger and Rabinowitz (1979) provided a somewhat different perspective on the relationship between class size and grades. They found that with larger classes faculty preferred to use a norm-referenced approach to grading, which usually yields lower grades than when factors such as effort are included. One explanation may be that larger classes do not lend themselves to high levels of faculty-student contact, thereby making it more difficult to evaluate each student's individual progress in the course. Another explanation may be that the larger classes tend to be introductory-level

courses or courses required of all students, while the smaller classes tend to be higher-level courses taken by more advanced students. Stein and Guthrie (1990) found a significant relationship between course level and mean GPA, with higher grades in the more advanced courses. They argued that the distribution of grades should be expected to differ from freshman courses to senior courses. Bogart and Kistler (1987) found that English composition instructors in both community colleges and state universities tended to apply and to give equal importance to the same criteria for grading in normal situations; this suggests that disciplinary differences transcend institutional type.

Faculty Grading Practices

A number of explanations have been offered for the variability of faculty grading practices. One argument is that professors have different philosophies regarding assessment that affect their grading practices. Dressel (1961) has identified three types of faculty members, based on their grading practices and philosophies: (1) traditionalists, who emphasize formal, objective measurement of knowledge and are critical of lax grading standards; (2) eclectics, who feel that, because the college experience is pluralistic, grading standards must recognize differing student backgrounds and interests; and (3) relativists, who view education as a continuing process of improvement and feel that grading interferes with actual learning. Etzioni (1975) has commented that declining standards may result when there is an attempt to combine these different approaches to grading into one system.

Geisinger and Rabinowitz (1979), building on the work of Thorndike (1969), identified three orientations toward grading: (1) norm-referenced, in which faculty evaluate students relative to the performance of others in the class; (2) criterion-referenced, in which students are evaluated in terms of previously set standards; and (3) self-referenced, in which faculty evaluate each student relative to his or her abilities, motivation, and/or past performance. In the early 1960s, according to Geisinger and Rabinowitz, norm-referenced assessment was common but there was a shift away from it in the middle to late 1960s toward other orientations, especially the self-referenced orientation. In the middle to late 1970s faculty appear to have favored the criterion-referenced orientation (Juola 1974; Carney, Isakson, and Ellsworth 1978; Bellott 1981). This was evident in the use of mastery models of learning that emphasized objective-based instruction, learning modules, and performance contracts. Geisinger and Rabinowitz found that faculty with a norm-referenced orientation gave lower grades than those with other approaches to grading. These findings suggest that rising GPAs may be the product of changes in faculty orientations toward grading. Faculty grading orientation appears to vary across types of institutions. In a study comparing attitudes at a university, a four-year college, and a two-year college, Geisinger and his colleagues (1980) found that university faculty were more norm-referenced than college faculty, and that community college faculty were most likely to advocate the self-referenced approach to grading. As Hanna and Cashin (1988) have pointed out, both norm-referenced (class curve) and criterion-referenced (percentage) grading systems lack a sound rationale; these authors recommend the use of anchor measures.

Raths, Wojtaszek-Healy, and Della-Piana (1987) concluded that there are two principal approaches to grading: (1) translating the grading problem into an arithmetic algorithm, giving points for subunits, weighing them according to the instructor's view of their importance, and establishing cutoffs for specific grades; and (2) using a more intuitive and less predictable approach, such as assigning grades based on a general impression of how well the work meets the instructor's expectations.

Ory and Ryan (1993) provided an excellent discussion of the advantages and disadvantages of using normative versus absolute grading methods. They argued that grading on the basis of improvement may be unfair; they consider grading on the curve, the distribution gap method, and percentage grading to be less defensible than either normative or absolute grading methods.

Faculty in different departments may prefer different grading orientations because they have different views about the purposes of grading. A survey of faculty at a state university (Hambleton and Murray 1977) found that grading was given the highest importance as a means of providing students with feedback on their course performance, followed in importance as a means of informing others about student performance and, then, as a means of motivating students to do good work. Faculty in the education department placed less importance on grading as a means of motivating students than faculty in other departments. Most faculty surveyed felt that pass/fail and pass/not recorded grading systems failed to meet the most important purposes of grading. Criterion-referenced grading was preferred over norm-referenced grading for meeting these purposes. These faculty believed that the most important course outcomes were to maximize the amount of learning, to maximize student enthusiasm, and to maximize student performance. They saw criterion-referenced grading as most likely to bring about these out-

comes. Similar studies have been done at other institutions. For example, Haughey (1977) found that the faculty at a state college believed that grades were an important motivator for students and that standards should be the chief criterion in determining grades.

Other researchers have theorized that faculty adapt their grading standards to the ability level of their students (Aiken 1963; Hills and Gladney 1968; Goldman, Schmidt, Hewitt, and Fisher 1974; Goldman and Slaughter 1976; Gallini 1982). That is, when student ability increases, faculty tend to use stricter grading standards and, when student ability decreases, grading standards are lowered. Adaptation-level theory has been used to relate this to differences in department grading standards. The argument has been made that lower-ability students migrate toward courses and fields that have lower grading standards. This, in turn, leads to higher expectations of students who take the more demanding courses as faculty adapt to the high ability of the class and grade more rigorously. Adaptation-level theory has also been used to explain grade inflation. For example, McKenzie and Tullock (1981) argued that as college admission standards drop and the quality of students decreases, faculty adapt by grading more leniently. Elliott (personal communication) has found that, at a variety of institutions, prospective science majors have higher mean scores on college admission tests than the average for all students entering the institution. This, along with adaptation theory, might explain the higher grading standards in science departments. Summerville and his colleagues (1988) concluded that departments do not award high grades because they attract high-achieving students. Their findings, that students in departments that traditionally assigned low grades received higher grades in their other courses, is consistent with adaptation theory.

Faculty may give greater or lesser emphasis to different grading criteria depending on the situation. Bogart and Kistler (1987) found no difference, overall, in the grading criteria used by English department faculty in community colleges and state universities nor in the importance they assigned these criteria in normal situations. In borderline situations, however, community college faculty gave significantly more importance to "final exam scores" and university faculty gave significantly more importance to "adherence to due dates and deadlines" and to "improvement shown since the beginning of the course."

Hughey and Harper (1983) found that a substantial portion of a course grade was the result of how the instructor differentiated among students. The communication responsiveness of the student (including things such as the kind of communication climate the student creates, the way the student transmits information, and the way the student deals with communication barriers) had a "pervasive influence" on final course grade. Females were more likely to get a high grade (gender accounted for 15.8 percent of the variance in grades).

Faculty characteristics, such as rank, seniority, tenure status, and age, may influence grading practices. Prather (1976) found that graduate teaching assistants and instructors gave higher grades, on average, than other undergraduate faculty. This suggests that the growing use of part-time faculty may be related to grade inflation. Boozer (1977) however concluded that grading patterns are idiosyncratic; he found no relationship between faculty rank, degree, tenure, and age and grades given. Dickson (1984) was also unable to support the hypothesis that junior faculty mark more easily.

In addition to varying at any one point in time, it is likely that faculty grading practices also vary over time. This is suggested by the increases in GPA observed since the late 1960s. These increases remain substantial even after controlling for the academic preparation of the students (Goldman, Schmidt, Hewitt, and Fisher 1974; Goldman and Slaughter 1976; Summerville, Ridley, and Maris 1988). Faculty may have experienced changes in attitudes or grading philosophies during the past two decades. After all, during this period the sorting function of education has come under severe attack (Rist 1970; Bowles and Gintis 1976; Oakes 1985). Faculty may have become increasingly sensitive to the role grades play in the sorting process and, as a result, responded by grading more leniently. Also, the introduction of student course evaluations and the greater use of part-time faculty during the past two decades may have contributed to grade inflation. It is also possible that faculty have modified their grading practices as a response to changes in the student body, including changes in academic preparation as well as changes in sex, race/ethnicity, and age.

Changes in Student Achievement Level and in the Perceived Importance of Grades

Grades may have risen because students now achieve at higher levels than did students in the past or because students are now more grade-oriented, thus making faculty feel they should award higher grades. Johnson and Beck (1988) found that students with low SAT scores are more grade-oriented than students with higher SAT scores.

In a survey by the Carnegie Foundation for the Advancement of Teaching (1989), 70 percent of faculty

agreed with the statement "Undergraduates have become more grade conscious." However, 55 percent of faculty also agreed with the statement "Most undergraduates at my institution only do enough to get by." Agreement with the second statement was higher at comprehensive colleges and universities (defined as institutions in which more than half the baccalaureates are given in occupational or professional disciplines) than at research universities, doctoral universities, or liberal arts colleges, but there was essentially no variation in agreement levels for the first statement. This suggests that faculty in comprehensive colleges and universities may feel pressured to give good grades to students who are performing at only marginal levels of achievement, possibly because occupational and professional course grades have a more direct impact on career outcomes than is the case for liberal arts course grades.

Students also respond differently to strict or lenient grading standards. Johnson and Beck (1988) found that students with low SAT scores did better when they were graded on a strict scale and that grading scale variations had their greatest influence on students with low SATs. Still other researchers, such as Sarafina and DiMattia (1978), have argued that grades are undesirable because they reduce students' intrinsic motivation.

The relationship of changing student achievement levels, changing orientation to grades, and reactions to grading standards merits further research.

In sum, the literature on grades and grading practices suggests that the observed differences across institutions and departments, as well as the rise in GPAs that has been reported at many institutions, especially during the 1970s, may have been the result of changing institutional or departmental policies and/or practices or of changing attitudes and behaviors among faculty. From the institutional perspective, GPAs may be related to admission standards, curricular requirements, grade reporting systems, the use of part-time faculty, the use of student course evaluations, and administrative pressures to limit grade inflation or to present the appearance of a uniform grading system. At the departmental level, many of the same factors apply. In addition, departmental differences in grading may be related to differing views of knowledge and approaches to assessment across disciplines, as well as to variations in class size and the level of the courses offered. The grades of individual faculty are probably also affected by their differing views of knowledge, the kinds of assessment they use, and their views about the purposes of assessment. Faculty may respond differentially to pressures from students to award higher grades. There is some evidence to indicate that part-time faculty, graduate assistants, and instructors grade less strictly than full-time faculty.

Although the literature on grades and grading practices in higher education is extensive, it nonetheless contains many gaps. For example, we do not know the extent to which differences in grading across disciplines result from differences in departmental policies, faculty attitudes, or differences in the types of students enrolled. Nor do we know if or how faculty grading philosophies are related to grading practices. In addition, there are dangers in trying to reach conclusions about current grading practices on the basis of the research summarized here because some of this research may be outdated; there were relatively few empirical studies of grading during the 1980s. As a result, we know little about current grading practices in higher education, how these may have changed during the past decade, and why. This study is a first step toward addressing some of these topics.

Method

This study sought information to answer four questions:

- What are the current institutional and/or departmental policies and practices related to grading?
- Were there changes in these policies and practices between 1980 and 1990? If so, what was the nature of these changes?
- What are the current grading practices and attitudes of college faculty who teach undergraduate courses? Have these changed over time?
- Do faculty grading orientation and practices differ across departments?

Instruments

Based on the preceding review of the literature and on suggestions from researchers who have previously studied college grading practices, three questionnaires were developed: (1) an institutional questionnaire, (2) a departmental questionnaire, and (3) a faculty questionnaire. The rationale for this three-questionnaire approach was that grades are likely to reflect institutional and departmental policies and expectations as well as individual faculty members' attitudes and grading approaches.

The institutional and departmental questionnaires were designed to ascertain if there were pressures on faculty that might affect their grading practices. The institutional questionnaire asked about the nature of the grade reporting system, policies related to grading, re-

quired courses and other curriculum matters, and institutional use of student evaluations, as well as for information about institutional characteristics.

The departmental questionnaire focused on departmental grading policies and/or expected grading practices, how these are communicated to faculty members, the extent of review of grades by department chairs, and the frequency of departmental discussions about grading. The rationale was to determine the extent to which departments try to develop some consensus about grading.

The faculty questionnaire was the most complex instrument of the three. It began by asking faculty about their institution and department, including their perception of student ability, in order to determine if there was a relationship between grading practices and perceived student characteristics. Faculty were then asked about their perceptions of the pervasiveness of departmental policies about grading to determine how much these might affect grading practices, and their perceptions of departmental efforts to raise grading standards, which might help to inform changing relationships between test scores and grades. The next section of the faculty questionnaire dealt with attitudes about grading, changes in grading philosophy, and opinions about the purposes of grading. This section also included questions about faculty perceptions of variations in the meaning of course grades, both across institutions and across disciplines, and faculty perceptions as to why undergraduate grades now tend to be higher than they were 20 years ago.

Questions about attitudes toward grading were included so the findings of this study could be related to Dressel's (1976) and Etzioni's (1975) theories about the relationship between grading philosophies and grading practices. Questions about the purposes of grading were included to permit exploration of Hambleton and Murray's (1977) finding that faculty in different departments hold differing views on this topic. A question about changes in grading philosophy was included to provide some information about the extent to which grade inflation may be related to changing faculty views rather than to institutional and/or departmental factors. A question about the reasons for grade inflation was included to determine if faculty felt the changes were related to changes in the institution, in the students, or among faculty.

The next section of the faculty questionnaire dealt with grading practices, including grading orientation (criterion-referenced, norm-referenced, or self-referenced, as described by Geisinger and Rabinowitz 1979). To obtain a better understanding of the factors that faculty consider in assigning grades and of how information about student learning is obtained, this section also asked questions about the importance of various factors in grading lower-division, upper-division, and field experience/internship courses, as well as the types of assessment used. As noted previously, other researchers have found that the factors included in a grade and the kinds of assessment vary, both by department and by course characteristics. A brief section on student evaluations of faculty was included to determine the extent to which faculty feel pressured by this practice. Finally, a series of background information questions (academic rank, years in teaching, tenure status, degree, age, sex, and race/ethnicity) were included.

Institutional Sample

The original sampling plan was to select 12 institutions, dividing the sample evenly between public and private and between less selective and more selective institutions. The rationale for the sampling plan was based on findings that SAT validity has tended to decline more in public colleges than in private colleges and more in institutions where the average entering student's SAT score (verbal and math combined) was below 950 than in institutions where the average entering student's SAT score was above 950 (Morgan 1989). One hypothesis about this decline in validity is that it might be related to changes in college grading practices. For this reason, the sample included some institutions that had participated in a previous study of the predictive validity of the SAT (Ramist, Lewis, and McCamley 1990). We also wished to include in this study institutions that had been the subjects of previous research on grading, so previous findings might help inform our own results. The institutions were not intended to be a representative sample and, therefore, no generalizations to other colleges and universities should be made.

The final sample consisted of 14 institutions. Of these, eight (57 percent) are public and six (43 percent) are private. Using the Carnegie classifications, three are research institutions, three are doctoral institutions, five are comprehensive institutions, two are liberal arts institutions, and one is a specialized institution. Three of the 14 institutions reported that the highest degree they award is the bachelor's; 5 award the master's; and 6 award a doctorate. Only 3 of the colleges and universities in this study had participated in the previous validity study, and they are all private institutions that would be rated "more selective" in admission.

The institutions ranged in size from approximately 2,000 to over 22,000 full-time undergraduate enrollments. Part-time enrollments ranged from under 100 to approximately 6,000. Average total undergraduate enrollment (both full time and part time) was 18,322 at

the research institutions, 7,691 at the doctoral institutions, 4,769 at the comprehensive institutions, and 3,108 at the liberal arts institutions. Part-time students represented approximately 15 percent of the total undergraduate enrollment at the research institutions, 38 percent at the doctoral institutions, 34 percent at the comprehensive institutions, and 1 percent at the liberal arts institutions.

The 14 institutions vary considerably in selectivity. According to the *College Handbook*, which provides data on average SAT scores for the middle 50 percent of entering students, the combined SAT (verbal plus mathematical) score at three of the institutions was below 950; these institutions are designated "less selective" in this analysis. At six, the average combined SAT score for the middle 50 percent of entering students ranged between 950 and 1100; these are designated "moderately selective." At four of the institutions, the average combined SAT score was above 1100 for the middle 50 percent of entering students; these are designated "more selective." No data on selectivity were available for the specialized institution, so it was excluded from all analyses involving selectivity; the best estimate is that it is in the more selective category. (It described itself as very selective when responding to the institutional questionnaire.) When asked to rate the selectivity of their current (1990) admission policies, one institution described itself as having open admission, seven said they were moderately selective, and six said they were very selective. When asked about their admission policies in 1980, three said they had open admission at that time, six said they were moderately selective, and five said they were very selective. Comparisons of institutional self-reports of current selectivity and selectivity as reported in the *College Handbook* are shown in Table 1; agreement was high.

The public institutions in this sample were more likely to award the doctorate than the private but they also appeared to have somewhat less selective admission policies. Public institutions were much more likely than private ones to have a Carnegie classification of comprehensive, indicating that at least half their bachelors degrees were awarded in applied areas. There were no public liberal arts colleges in this sample.

TABLE 1

Institutional Selectivity in 1990: Self-Reported Versus *College Handbook*

College Handbook Estimate	*Self-Reported*		
	Open	*Moderately Selective*	*Very Selective*
Less selective	1	2	0
Moderately selective	0	6	0
More selective	0	0	4*

* Does not include the specialized institution estimated to be more selective but for which no *College Handbook* data was available.

Department and Faculty Sample

In order to obtain information about grading practices in a variety of academic areas, seven departments and the faculty members in them were selected for study, given that the institutions had such departments. In selecting departments, the rationale was to include those in which freshman students are likely to enroll; to include the sciences, social sciences, and humanities; and to include departments with an occupational orientation as well as departments in the liberal arts. Efforts were made to select departments that were reported to vary in their grading standards or orientation. The departments selected were: business, chemistry, education, English, history, mathematics, and psychology. All faculty members in each of the selected departments were included in the faculty sample.

Data Collection

Each institution that agreed to participate in this study was sent a copy of the institutional questionnaire. It had the option of completing and returning the questionnaire or of answering the questions during a telephone interview; most chose the latter but also provided supplementary written materials.

The institutions were asked if they had any of seven targeted departments and, if so, to provide the name of the department chairperson and a roster of the faculty. All 14 colleges and universities had undergraduate departments in chemistry, English, history, and mathematics. All but one had a psychology department. Seven had an undergraduate business department and nine had an undergraduate education department.

Each of the 85 department chairpersons was sent a letter explaining the study, a copy of the departmental questionnaire, and a copy of the faculty questionnaire. The department chairs were asked to complete both a departmental and a faculty questionnaire and return them to Educational Testing Service (ETS). A month later, each received a follow-up postcard asking them to return their questionnaires if they had not already done so and to encourage faculty in their department to return the questionnaires. Completed questionnaires were returned by 68 percent of the department chairs (58 individuals).

Faculty questionnaires were sent to all individuals on the rosters of smaller departments and to all cur-

rently active full-time faculty (to the extent that part-time, adjunct, and emeritus faculty could be determined) in the larger departments. Therefore these findings may underrepresent differences between full-time and part-time faculty. In most cases arrangements were made for the faculty questionnaires to be returned directly to ETS but, in a few cases, departmental cooperation was contingent on the questionnaires being returned to the department chairs and then forwarded to ETS. Anonymity of the respondents was important and it was not possible to do individual follow-ups with faculty who did not return their questionnaires. The total number of faculty questionnaires distributed was 2,180. Completed usable questionnaires were returned by a disappointing 25 percent of the faculty (542 individuals). Return rates varied considerably across institutions.

Results

Institutional Analysis

Curriculum

As discussed previously, curriculum changes occurred at many colleges and universities in the period between 1980 and 1990. The 14 responding institutions were no exception. Overall, the picture showed a move toward greater specification of curriculum. While none of these institutions had put a core curriculum in place during the decade, seven had instituted or added general education courses that all students were required to take and one had instituted such courses for specified students. The newly required courses included history, writing, non-Western culture, and foreign language. Four of the institutions adding new course requirements are privately controlled and more selective; three are public and rate themselves as moderately or very selective. An additional four of the responding institutions had added or instituted distribution requirements whereby students had to choose courses in specified fields or areas. None of the institutions reported having eliminated either a core curriculum or required courses, but one reported eliminating distribution requirements.

Three of the 14 institutions said they had made no change between 1980 and 1990 in the number and type of courses required of undergraduates, but one of these reported that plans for changes were in progress. The two institutions that had made no curriculum changes and had no immediate plans to do so are both public institutions that award doctoral degrees; one is less selective and one is moderately selective. Both these findings and those from other recent surveys (Lewis and Farris 1989; Toombs and Fairweather 1989) suggest that changes in general education requirements between 1980 and 1990, such as the addition of required courses for all students, may have affected college grades, especially in the freshman year. Narrowing students' course choices makes it more difficult for them to avoid courses in which they believe they will not do well. Also, requiring a course is often associated with larger class size, which may also have affected grades. Both of these factors would be expected to result in lower grades.

Grading Systems

The colleges and universities responding to the survey were asked to indicate all grading systems currently used and, also, all systems used in 1980. The results, shown in Table 2, suggest that these institutions moved toward a more differentiated grading system than they had in 1980. One institution had used descriptive evaluations in 1980 but had abandoned this system by 1990.

Twelve of the 14 colleges and universities used pass/fail and/or credit/no credit grading in 1990. Among these there was a tendency to place some limitations on this type of grading. Two-thirds limited the number of courses students could take under pass/fail or credit/no credit grading. Most (83 percent) did not allow students to receive pass/fail or credit/no credit grades in their major. And in most of these institutions relatively few students (estimates generally ranged from 1 percent to less than 5 percent) received credit/no credit or pass/fail grades; however, one university reported that approximately 20 percent of its students received this type of grade.

Each institution was asked how it defined an "average" grade in an undergraduate course. Seven of the institutions indicated that an average undergraduate grade was defined as C, equivalent to 2.0. Five replied that there was no official definition; of these one indicated that C was accepted as average performance while another said that empirically the average was B–. One

TABLE 2

Number of Institutions Using Various Grading Systems in 1990 and in 1980

	1990	*1980*
Five-letter system (A,B,C,D,F)	7	9
Letter system with plus and minus	8	4
Pass/fail	10	11
Credit/no credit	7	7
Descriptive evaluations	0	1

institution said the average undergraduate grade, defined on a bell-shaped curve, was 2.62 or C+. One said that it defined the average undergraduate grade as 2.7.

While a majority of the 14 institutions said they had no unusual grading policies or practices, five reported policies that they believed to be unusual. These included: (1) not giving grades of A+; (2) differing grading policies among the university's undergraduate schools and colleges; (3) rules regarding recording of pass/fail grades, including whether or not the transcript indicates there is a nonrecorded grade; (4) changing from having no final exams and no grades, except narrative evaluations, in 1980, to using traditional letter grades almost exclusively in 1990; and (5) waging a campaign from 1982 to 1990 to "stem grade inflation," which was reported to have reduced the overall undergraduate GPA from 3.0 to 2.7.

Institutional Grading Policies

The responding institutions were asked about their current (at the time of questionnaire completion in 1990) and 1980 policies in five areas related to grading: (1) incompletes and withdrawals; (2) recording grades; (3) computing the GPA; (4) other grading policies; and (5) GPA requirements. These are summarized in Table 3.

As Table 3 indicates, grading policies at these institutions remained fairly stable between 1980 and 1990. The overall picture is one of fairly tight institutional control over time limits for withdrawals from courses. All 14 institutions placed some restrictions on withdrawal from a course. In addition, passing grades were required for course withdrawal in nearly half the institutions. Only one allowed students to withdraw from a course at any time and without penalty. With regard to making up incompletes, all the institutions had a time limit in 1990 and all but one had a limit in 1980.

Institutional control over grades was also evident; at most institutions students had to meet minimum GPA requirements to remain in good standing and to graduate. In addition, at over half the institutions students were placed on probation when their GPA fell below the stipulated minimum. At 6 of the 14 institutions, the policy of using both grades from repeated courses when computing the GPA established some control over grade inflation. But policies pertaining to how the GPA is computed that might increase GPA inflation were noted as well. These included using only the last grade from repeated courses and using grades from remedial courses when computing the GPA.

Student Evaluation of Faculty

Of the 12 institutions responding to the question about student evaluation of instructors, 11 reported having

TABLE 3

Summary of Institutional Grading Policies in 1990 and in 1980

	1990	*1980*
A. Incompletes and withdrawals		
Time limit for making up incompletes	14	13
Withdrawal at any time without penalty	1	1
Withdrawal only in certain time period	12	12
Passing required for withdrawal	0	0
Withdrawal only in certain time period and if passing	6	6
B. Recording grades on transcript		
Second grade replaces first when repeating a course	3	3
Failing grades not recorded	1	1
D's not recorded	0	0
C. Computing the GPA		
Use only grades from courses that count toward degree	3	2
Use both grades from repeated courses	6	6
Use last grade from repeated courses	5	5
Include grades in remedial courses	4	4
D. Grading policies (not usual practices)		
Course average must be between B and C	0	0
Must grade against specific standards	3	3
Must grade relative to overall class	0	0
Must grade relative to student's ability	1	1
Must grade on progress toward individual objectives	0	0
Must include effort in grade	0	0
Must include attendance in grade	1	1
Must grade on curve	0	0
E. GPA requirements		
Minimum GPA required for good standing	12	12
Probation when GPA falls below minimum	12	12
Probation when any semester below minimum	6	5
Students on probation suspended when semester GPA is below minimum	7	7
Students remain on probation until overall GPA rises above minimum	8	9
Minimum GPA required for graduation	12	12
Minimum GPA in major required for graduation	8	8

such evaluations. Among these institutions, 11 used student evaluations as part of faculty retention, promotion, and tenure decisions in 1990 and 8 used student evaluations in such decisions in 1980; the increase occurred in public, moderately selective institutions. One institution allowed faculty to determine whether or not student evaluations would be used in making promotion and tenure decisions. Five institutions said they used student evaluations as part of faculty merit pay decisions in 1990, while only two made this use of student

TABLE 4

Institutional Uses of Student Evaluations of Faculty in 1990 and in 1980 (N=11)

	1990	*1980*
Faculty retention decisions	11	8
Faculty promotion decisions	11	8
Tenure decisions	11	8
Merit pay decisions	5	2

evaluations in 1980. This increase occurred in less and moderately selective public institutions. Four reported using student evaluations to make decisions about teaching awards. Three institutions said they used student evaluations to help faculty modify a course or improve teaching methods. These data are summarized in Table 4.

Department Chairperson Analysis

Questionnaires were received from 59 department chairpersons; 3 from business, 11 from chemistry, 5 from education, 8 from English, 12 from history, 7 from mathematics, 9 from psychology, and 4 for whom departmental identification was not available. Slightly more than half (55 percent) of the department chair responses came from public institutions and 45 percent from private institutions, closely reflecting the representation in the institutional sample. About a third of the responses were from comprehensive institutions, about 22 percent from research institutions, another 22 percent from doctoral institutions, and about 20 percent from liberal arts institutions. About 20 percent of the department chair responses came from less selective, about 43 percent from moderately selective, and about 37 percent from more selective institutions.

None of the department chairs reported having specific departmental grading policies and only one department, the history department at one of the more selective private institutions, reported having a specific expected grading practice. (Expected practices are not formally stated, but are transmitted to department members through informal means.) Thus it appears that departmental grading policies and explicitly stated departmental grading expectations played a small role in the grading process.

Sixty-two percent of the department chairs responding to the questionnaire said they routinely reviewed the grades faculty gave in courses in their departments. Variations by department and by institutional characteristics are shown in Table 5. It is not known whether or not faculty grading practices were affected by knowledge that the grades would be reviewed by the department chair.

TABLE 5

Percentage of Department Chairs Reviewing Faculty Grades

A. By department	
Education	100
Mathematics	86
English	75
Business	67
History	64
Chemistry	55
Psychology	33
B. By institutional selectivity	
Less	100
Moderate	52
More	29
C. By Carnegie category	
Comprehensive	76
Doctoral	64
Liberal arts	44
Research	18
D. By institutional control	
Public	64
Private	45

Although many department chairpersons evidenced a concern about grading practices through their routine reviews of the grades given, only about a third said they ever had formal meetings of department faculty to discuss grades. Of these, about 80 percent had one or two and 20 percent had three to five such discussions per year. The frequency of such discussions, by department and by selected institutional characteristics, is shown in Table 6. It seems likely that faculty in departments that have formal discussions of grading will have a more uniform view of grading standards and practices than faculty in departments that do not have such discussions.

Informal discussions of grades were more common and were reported by approximately 75 percent of the department chairs. Half of the department chairpersons reported having informal discussions about grading with faculty one or two times a year, while another 32 percent reported having such discussions three to five times a year and the remaining 18 percent reported informal discussions about grading from once a month to once a week. The frequency of these discussions, by department and by institutional characteristics, is shown in Table 6.

The evidence from this portion of the analysis indicates that departmental influences on grading work primarily through informal and formal discussions about

TABLE 6

Percentage of Department Chairs Reporting Formal and Informal Discussions of Grading, by Departmental and Institutional Characteristics

	Formal	*Informal*
A. By department		
Education	60	100
Chemistry	45	91
Psychology	44	89
History	33	67
Mathematics	14	71
English	12	63
Business	0	66
B. By institutional selectivity		
Less	40	100
Moderate	38	67
More	17	45
C. By Carnegie category		
Comprehensive	47	88
Research	36	45
Liberal Arts	20	80
Doctoral	9	73
D. By institutional control		
Public	39	78
Private	22	70

grading rather than through explicitly stated departmental grading policies or prescribed practices. The responses suggest that departmental influences on grading are most likely to be found in less selective colleges and universities and in education, chemistry, and psychology departments. The rather limited amount of discussion about grading reported in English departments implies that grading practices may vary more in this field than in others. Discussions about grading were reported most frequently by department chairs in less selective institutions and in comprehensive institutions. This suggests that department chairs in institutions that enroll students more diverse in ability and goals may see a greater need to help faculty develop a consensus on grading than do department chairs in institutions in which students are more similar.

Individual Faculty Member Analysis

This section reports the results from questionnaires completed by 597 faculty members. Department chairs with teaching responsibilities also completed a faculty questionnaire.

TABLE 7

Number of Faculty Responding by Field and Institutional Type (N = 597)

	Public	*Private*	*Unknown*	*Total*	*Percent of Sample*
Business	40	16	13	69	10.1
Chemistry	73	30	7	110	18.4
Education	38	13	1	52	8.7
English	76	36	2	114	19.1
History	35	39	4	78	13.1
Mathematics	72	14	8	94	15.7
Psychology	45	32	2	79	13.2
Field unknown	10	0	0	10	1.7

Table 7 shows the number and percent of faculty responding, by field and by institutional control. Faculty from English and chemistry departments are most frequently represented in the sample, while education department faculty are less frequently represented.

Table 8 summarizes faculty background information. Nearly half the responding faculty said they held the rank of full professor and about three-quarters had received tenure. Nearly all the respondents taught full time and nearly all held a doctoral degree. Slightly more than half the respondents had been teaching at the college level for 16 years or longer. About three-quarters were male and nearly all were white. More than half were under 50. Among those indicating their age, about 26 percent were 39 or younger, 34 percent 40 to 49, 27 percent 50 to 59, and 13 percent 60 or older.

Departmental Information

Because the characteristics of the departments in which they teach may be associated with faculty members' grading practices, information about the highest degree awarded, number of majors, teaching loads, and faculty perceptions of student quality was collected.

About 38 percent of the respondents taught in departments in which the highest degree awarded is the bachelor's, 24 percent in departments in which the highest degree is the master's, and 38 percent in departments that award the doctorate. Slightly more than half the respondents from psychology and history departments indicated their departments awarded the doctorate as the highest degree, but only 16 percent of the business faculty were from departments that award the doctorate. Slightly more than half the respondents from chemistry and mathematics departments said the

TABLE 8

Percentage of Responding Faculty with Various Background Characteristics ($N = 597$)

	Percent
A. Rank	
Full professor	45
Associate professor	28
Assistant professor	19
Instructor	7
B. Tenure status	
Tenured	75
Not tenured	25
C. Employment status	
Full time	97
Part time	3
D. Teaching experience	
16 or more years	55
8 to 15 years	26
4 to 7 years	11
0 to 3 years	8
E. Sex	
Male	76
Female	24
F. Race/ethnicity	
White	95
Other	5

highest degree awarded was the bachelor's, but only 19 percent of the responding psychology faculty indicated this was the highest degree in their department.

The literature on grading suggests that the level of the courses that faculty teach might be associated with grading practices. The faculty members in this study, on average, devoted about 47 percent of their teaching time to introductory/lower-division courses, 41 percent to advanced/upper-division courses, and 22 percent to graduate courses. (The percentages do not total 100 because some faculty taught no graduate courses.) Faculty in mathematics and in chemistry reported spending the highest proportion of their time teaching introductory courses (both 57 percent). All other faculty averaged less than 50 percent of their time teaching introductory courses, with education and psychology faculty reporting the least time teaching at this level.

The departments varied considerably in the number of undergraduate majors, ranging from less than 10 to more than 100. While nearly half the business departments reported having more than 100 majors, none of the chemistry departments and only 3 percent of the mathematics departments reported this many majors. Business, education, and psychology departments averaged between 76 and 100 undergraduate majors while English and history departments averaged between 51 and 75. Mathematics departments averaged between 26 and 50 majors while chemistry departments averaged 11 to 25.

The literature on grading suggests that both the number of classes taught and the size of these classes might be related to grading practices. Faculty members responding to the questionnaire reported teaching, on average, 2.2 undergraduate courses each term. Chemistry faculty reported teaching the fewest courses per term, with an average of 1.66, while English and business faculty reported teaching the most courses, 2.49 and 2.46 respectively. Responding faculty members indicated that their introductory/lower-division classes averaged about 55 students and their advanced/upper-division classes about 19 students. Introductory classes tended to be much smaller in English (averaging 34 students), and in mathematics, education, and business (all averaging 38 students) than in chemistry (89 students), history (71 students), and psychology (66 students). There was less variation reported in the size of advanced/upper-division classes, with the smallest classes in mathematics (18 students) and the largest in psychology (36 students). The average advanced/ upper-division class size reported was 26 students.

The literature on grading also suggests that faculty may adjust their grades to the ability level of their students. In this study, most faculty (69 percent) felt that the students in their department were as well prepared as students in other departments at their institution; 29 percent felt their students were better prepared, and 2 percent felt their students were less well prepared. Faculty in chemistry, education, and business departments were most likely to feel that their students were better prepared than the average student in their institution.

Efforts to Raise Departmental Standards

Some believe that grading practices have changed in recent years in response to departmental efforts to raise standards. In this study, about half (56 percent) of the respondents said their departments had tried to raise standards during the previous 10 years, 30 percent said their department had not tried to raise standards, and 14 percent did not know whether or not their department had tried to raise standards during this period (see Table 9). Only about a quarter of chemistry department faculty (26 percent) said their departments had tried to raise standards, while about 85 percent of education department faculty reported such efforts. Approximately 20 percent of chemistry and mathematics faculty said they did not know if there had been departmental ef-

TABLE 9

Percentage of Responding Faculty Reporting Department Efforts to Raise Standards, by Field (N = 558)

	Yes	*No*	*Don't Know*
Business	73	14	13
Chemistry	26	55	19
Education	85	4	11
English	60	27	13
History	53	41	6
Mathematics	52	28	20
Psychology	66	23	10
Total	56	30	14

forts to raise standards, while only 6 percent of history faculty indicated they did not know if there had been such efforts.

Departmental Grading Policies and Expected Practices

As indicated earlier, we believed that grading might be influenced by institutional and/or departmental policies as well as by individual faculty members' grading philosophies and practices. In order to determine the kinds of grading behaviors that faculty believed the department required or expected, faculty were asked if certain grading practices were required by departmental policy or were expected departmental practice. The results are summarized in Table 10.

Few faculty reported that there were departmental grading policies, supporting the evidence from the departmental questionnaires. The most frequent policy, grading students against specific standards, was reported by about a quarter of all faculty. Other grading policies most frequently mentioned by responding faculty were giving students an A or a B in an honors course, including attendance in the course grade, and grading students in relation to the overall performance of the class.

Departmental expectations about grading were more commonly reported than were explicit grading policies. Again, grading against specific standards was mentioned most frequently (by nearly two-thirds of the faculty). Grading students relative to overall class performance was described as an expected practice by about half the respondents as was giving a lower grade on work turned in late. Slightly less than a third of the respondents said their departments expected them to give an A or B to students in an honors course, to include attendance in the course grade, to include effort in the course grade, or to grade students on how well they progressed toward individual objectives.

TABLE 10

Percentage of Responding Faculty Reporting Various Departmental Policies and Expected Practices Related to Grading

	Official Policy	*Expected Practice*
Students are graded against specific standards	26	63
Students are graded relative to overall performance of class	10	51
Students are graded on how well they progress toward individual objectives	9	30
Students' effort to learn is included in their course grade	3	30
Students' attitude and/or behavior is included in their course grade	6	26
Students' attendance is included in their course grade	11	31
Students are graded "on the curve"	5	27
The average course grade must be between a B (80) and a C (70)	6	22
Students should receive an A or a B in an honors course	14	32
Students should receive an A or a B in a fieldwork/internship course	3	21
Students can do extra credit projects to raise a grade	4	15
When students turn in work late, it must be graded lower	7	47
When students miss several classes, they receive a lower grade	5	22

Faculty Attitudes about Grading

Faculty were asked about their grading attitudes and philosophy, whether their grading philosophy had changed since they began teaching, the importance they attached to various purposes of grading, their beliefs about the variations in the meaning of grades, and the factors they believed most influenced the higher grades received by current students.

To assess grading attitudes and philosophy, faculty members were asked to indicate the extent to which they agreed or disagreed with seven statements (shown in Table 11). As Table 11 indicates, responding faculty were most likely to agree with the statement "I view education as a continuing process of improvement" ($M = 3.53$); there was also a high level of agreement with the statement "I emphasize formal, objective measurement of knowledge" ($M = 3.01$). Faculty least agreed with the statement "I consider grading as a harmful exercise which interferes with actual learning" ($M = 1.83$); the statement "Grading standards must recognize differing student backgrounds and interests" ($M = 2.00$) also received relatively low endorsement.

Variations in grading philosophies by department are also shown in Table 11. In this and most other ta-

TABLE 11

Faculty Attitudes about Grading, Means by Field and for Total Group, and Probability of *F*-test Ratio for Significant Differences Across Fields

	Business	*Chemistry*	*Education*	*English*	*History*	*Mathematics*	*Psychology*	*Total Group*
I emphasize formal, objective measurement of knowledge	3.13	3.41	2.65	2.47	2.57	3.37	3.35	3.01 p=.000
Grading standards must recognize differing student backgrounds and interests	1.98	1.77	2.35	2.37	1.99	1.74	1.90	2.00 p=.000
I am critical of nonobjective grading practices	2.29	2.93	2.10	2.23	2.15	2.90	2.45	2.49 p=.000
Grades cannot be reduced to a set of formal, objective measures	2.47	2.27	2.94	2.96	3.04	2.28	2.22	2.57 p=.000
Most faculty grade too leniently	2.60	2.68	2.51	2.66	2.71	2.56	2.77	2.65 p= ns
I view education as a continuing process of improvement	3.58	3.53	3.63	3.52	3.52	3.44	3.53	3.53 p= ns
I consider grading as a harmful exercise which interferes with actual learning	1.82	1.62	2.22	2.11	1.55	1.81	1.74	1.83 p=.000

Scale: 1 = strongly disagree, 2 = disagree, 3 = agree, 4 = strongly agree.

bles where differences by department are presented, analysis of variance was used to determine if there were significant differences among the groups. There were no significant differences across fields in views about the leniency of grading (F = 1.1875, df = 559, p = .3112) or about education as a continuing process of improvement (F = .8513, df = 571, p = .5307). Examination of the other five statements reveals that the departments clustered into two major groups, one seeing grades as formal and objective, and the other believing that grades cannot be reduced to a set of objective measures. Rather predictably, the "objective" camp includes chemistry, mathematics, and psychology departments while the "nonobjective" camp includes English, education, and history departments. Faculty in business appeared to be the most diverse in their attitudes; they could not easily be classified into either of these two groups.

There were significant differences across fields in faculty endorsement of the statement "I emphasize formal, objective measurement of knowledge" (F = 26.4582, df = 576, p = .0000), with English and history faculty differing significantly from business, chemistry, mathematics, and psychology faculty. Education faculty differed significantly from chemistry, mathematics, and psychology faculty on this item. There were also significant differences in endorsement of the statement "Grading standards must recognize differing student backgrounds and interests" (F = 9.8041, df = 581, p = .0000), with education faculty differing significantly from chemistry and mathematics faculty, and English faculty differing significantly from chemistry, mathematics, and psychology faculty. Faculty also differed significantly in their endorsement of the statement "I am critical of nonobjective grading practices" (F = 13.2648, df = 569, p = .0000), with chemistry faculty differing significantly from business, education, English, history, and psychology faculty, and mathematics faculty differing significantly from business, education, English, and history faculty. Endorsement of the statement "Grades cannot be reduced to a set of formal, objective measures" (F = 17.3849, df = 567, p = .0000) showed chemistry, mathematics, and psychology faculty differing significantly from education, English, and history faculty, and business faculty differing significantly from English and history faculty. Endorsement of the statement "I consider grading as a harmful exercise which interferes with actual learning" (F = 7.6784, df = 576, p = .0000) showed education and English faculty differing significantly from chemistry and history faculty.

Forty-three percent of responding faculty said they had changed their grading philosophy since they began teaching. As shown in Table 12, this percentage varied across departments, ranging from a high of 58 percent among education department faculty to a low of 33 percent among chemistry faculty.

Faculty were next asked to indicate the importance they assigned to each of 10 possible purposes of grading. The results are shown in Table 13. Faculty assigned highest importance to providing students with feedback (M = 3.64), providing other education institutions such as graduate or professional schools with in-

TABLE 12

Percentage of Faculty Who Had Changed Their Grading Philosophy since They Began Teaching

Business	47
Chemistry	33
Education	58
English	50
History	39
Mathematics	41
Psychology	43
Total	43

formation about students (M = 3.21), motivating students to do good work (M = 3.13), and helping the college or department make decisions about students (M = 3.05). There were no significant differences across fields in the importance assigned to grades as a source of information for student decisions (F = .9045, df = 582, p = .4912) or information for the college to make decisions (F = .7364, df = 581, p = .6205). There was a significant overall F for differences across fields in the importance of grades to provide feedback to students (F = 2.8697, df = 585, p = .0092), information to other institutions (F = 2.56, df = 584, p = .0187), and information to employers (F = 2.8239, df = 584, p = .0102) but, for all three items, the Scheffe procedure indicated no statistically significant differences between any pairs of fields.

There were significant differences across fields in the importance of grades as a source of instructor feedback about teaching effectiveness (F = 3.1386, df = 584, p = .0049), with education faculty differing significantly from English faculty. There were also significant differences across fields in the importance of grades in motivating students (F = 5.1795, df = 581, p = .0000), with education faculty differing significantly from faculty in chemistry, history, and mathematics. Faculty also differed significantly in the importance of grades as a way of preparing students for the competitive nature of adult life (F = 5.2400, df = 579, p = .0000), with education faculty differing significantly from chemistry and history faculty; in the importance of grades to help students learn discipline for work or jobs (F = 6.7850, df = 583, p = .0000), with education faculty differing significantly from business, chemistry, English, history, and mathematics faculty; and in the impor-

TABLE 13

Faculty Opinions about Purposes of Grading, Means by Field and for Total Group, and Probability of F-test Ratio for Significant Differences Across Fields

	Business	*Chemistry*	*Education*	*English*	*History*	*Mathematics*	*Psychology*	*Total Group*
Provide students with feedback	3.73	3.72	3.54	3.45	3.64	3.68	3.75	3.64 p=.001
Provide students with information for making educational and vocational decisions	3.00	3.04	2.98	2.88	2.91	3.12	2.96	2.98 p=ns
Provide instructor with information about teaching effectiveness	2.87	2.72	3.17	2.62	2.91	2.78	2.94	2.82 p=.005
Help college/department make decisions about students	2.93	3.06	3.02	3.06	3.11	3.17	3.01	3.05 p=ns
Provide other educational institutions with information about students	2.95	3.28	3.04	3.18	3.32	3.26	3.32	3.21 p=.019
Provide potential employers with information about students	2.75	3.00	2.92	2.68	2.86	2.95	2.62	2.82 p=.010
Motivate students to do good work	3.19	3.19	2.71	3.14	3.27	3.30	2.94	3.13 p=.000
Help prepare students for the competitive nature of adult life	2.58	2.68	1.98	2.48	2.82	2.53	2.34	2.52 p=.000
Help students learn discipline for later work or jobs	2.85	2.95	2.15	2.71	3.01	2.75	2.54	2.74 p=.000
Help maintain academic standards	2.97	3.02	2.62	2.89	3.19	3.14	2.91	2.98 p=.004

Scale: 1 = no importance, 2 = slight importance, 3 = moderate importance, 4 = great importance.

tance of grades to help maintain academic standards ($F = 3.1824$, $df = 582$, $p = .0044$), with education faculty differing significantly from faculty in history.

Education faculty stand out as differing most from their colleagues in other fields. They are more likely to see grades as providing instructors with information about their teaching effectiveness and are less likely to value grades as a way of motivating students, to help students prepare for the competition of adult life, or to help maintain academic standards.

Next faculty were asked how much they thought the meaning of a course grade varies: (a) across disciplines in their institution and (b) within their discipline across institutions. The scale used was 0 = none, 1 = a little, 2 = some, and 3 = a lot. Faculty perceived greater variation in grading standards across disciplines ($M = 2.60$) than within a given discipline ($M = 2.15$). The level of perceived differences in the meaning of course grades across disciplines did not vary significantly by field ($F = 1.0080$, $df = 535$, $p = .4192$). The level of perceived differences within disciplines varied significantly ($F = 4.1570$, $df = 524$, $p = .0004$) across fields, with chemistry differing significantly from psychology.

Faculty were asked why they thought grades are generally higher today than they were 20 years ago. From a list of 13 possible factors, they were asked to select *three* that they believed were important and then to indicate which *one* of these was most important. A quick glance at Table 14, which summarizes the findings, shows that there was considerable diversity of opinion. The reason selected most frequently, both as the single most important factor (by 30 percent of responding faculty) and as one of three most important factors (by 54 percent), was "Faculty expect less of students." In second place was "Faculty have less concern for maintaining high standards," selected as the most important reason by 9 percent of responding faculty and as one of the three most important factors by 36 percent. Other factors that responding faculty also frequently indicated as important included "A change in focus toward student improvement rather than absolute standards" and "Liberalized course withdrawal policies." The latter finding conflicts with the institutional data (see Table 4) that did not show any changes in course withdrawal policies.

TABLE 14

Percentage of Responding Faculty with Various Opinions about the Most Important Factors in Grade Inflation

	Most Important Reason	*One of the Three Most Important Reasons*
Faculty expect less of students	30	54
Faculty have less concern for maintaining high standards	9	36
Liberalized course withdrawal policies	7	31
A change in focus toward student improvement rather than absolute standards	8	25
Students today are more competitive for grades	5	21
Changes in student aspirations regarding academic work	6	19
Greater diversity of student populations	4	19
Overcrowding classes forcing a change in grading standards	3	16
Students repeating courses to raise their grades	2	12
Better academic preparation on the part of students	4	11
A more mature student body	2	9
Use of pass/fail grading systems	1	9
Minimum competency testing has raised postsecondary standards	1	6

Grading Practices

The next section of the questionnaire asked about faculty grading practices, including the grading system(s) used, how they thought their grades compared with those of other faculty, the importance of various elements in assigning grades, the types of assessment used, and the testing process.

Faculty were first asked if they ever used any or all of three grading approaches: (1) grading based on some absolute standard (criterion-referenced); (2) grading students in relation to the performance of other students (norm-referenced); and (3) grading students in relation to their own improvement (self-referenced).

As Table 15 shows, 81 percent of faculty said they sometimes used a criterion-referenced approach to grading, while 57 percent sometimes used a norm-referenced approach and 44 percent sometimes used a self-referenced approach. There was no difference across fields in the use of a criterion-referenced approach ($F = 1.4270$, $df = 569$, $p = .2019$). There were significant differences in the use of a norm-referenced approach to grading ($F = 3.3823$, $df = 559$, $p = .0029$), with Scheffe tests showing that business faculty used a norm-referenced approach significantly more frequently than faculty in education. There were also significant differences across fields in the use of a self-referenced approach to grading ($F = 9.3700$, $df = 540$, $p = .0000$), with education, English, and history faculty using this approach significantly more frequently than mathematics faculty and English faculty using it significantly more frequently than business and psychology faculty.

Next faculty were asked which *one* of these approaches they used *most* often. About 64 percent of re-

TABLE 15

Grading Pr... ...es of Responding Faculty and Probability of *F*-Test Ratio for Significant Differences Across Fields

	Percentage Yes							
	Business	*Chemistry*	*Education*	*English*	*History*	*Mathematics*	*Psychology*	*Total*
Students are graded on the way their achievement level compares with some absolute standard	73	84	73	79	83	87	84	81 *p*=ns
Students are graded in relation to the performance of other students	70	61	35	61	48	52	64	57 *p*=.003
Students are graded in relation to their own improvement	44	38	61	67	54	22	36	44 *p*=.000

sponding faculty indicated that they most often used a criterion-referenced approach. About 29 percent said they most often used a norm-referenced approach, and about 8 percent said they most often used a self-referenced approach. Variations by field are shown in Table 16. There were no statistically significant differences across fields in preference for the criterion-referenced approach ($F = 1.2812$, $df = 519$, $p = .2641$) or the norm-referenced approach ($F = 1.8822$, $df = 519$, $p = .0820$). However, there were significant differences in preference for the self-referenced approach ($F = 3.4304$, $df = 519$, $p = .0025$), with faculty in education, English, and history most likely to prefer this approach to grading. The Scheffe procedure showed no significant differences across disciplines.

Responding faculty members tended to believe they were somewhat stricter graders than were other faculty in their department, in other departments in their institution, and in their discipline in other institutions. When asked to rate their grades in introductory courses in comparison with the grades of others, using the scale 1 = lower, 2 = about the same, and 3 = higher, the typical response was slightly lower than others in their discipline in other institutions ($M = 1.92$) and others in their department ($M = 1.81$), and moderately lower than faculty in their institution in other departments ($M = 1.59$). Variations across fields are shown in Table 17. Education faculty were the only group to report that they graded somewhat higher than faculty in other departments of their institutions. Chemistry faculty were the only group to report that they graded somewhat higher than faculty in the same field in other institutions. There were no significant differences across fields in regard to how grades were believed to compare with other faculty in the same department ($F = 1.6769$, $df = 496$, $p = .1246$) or with faculty in the same discipline in other institutions ($F = .3761$, $df = 287$, $p = .8939$). There were significant differences when faculty compared their grading with faculty in other departments at their own institution ($F = 15.9316$, $df = 420$, $p = .0000$). There were also significant differences between chemistry faculty and faculty in business, education, English, and psychology; between mathematics faculty and faculty in business, education, and psychology, and between faculty in education and those in English and psychology.

Faculty were asked to indicate the importance they gave to various factors in grading introductory/lower-division courses and in grading advanced/upper-division courses. The results are summarized in Table 18. Tests and quizzes had the strongest influence on grades in introductory courses ($M = 3.61$) and somewhat less influence in advanced courses ($M = 3.38$), while papers and written assignments, which had

TABLE 16

Grading System Most Often Used by Responding Faculty and Probability of *F*-Test Ratio for Significant Differences Across Departments

	Percentage of Faculty							
	Business	*Chemistry*	*Education*	*English*	*History*	*Mathematics*	*Psychology*	*Total*
Absolute standard	54	60	65	59	66	73	68	64 *p*=ns
Relative to other students	44	34	20	28	23	26	28	29 *p*=ns
Relative to individual improvement	2	6	15	14	11	1	4	8 *p*=.003

TABLE 17

Faculty Perceptions of How Grades in Their Introductory Courses Compared with Those of Other Faculty

	Business	*Chemistry*	*Education*	*English*	*History*	*Mathematics*	*Psychology*	*Total*
In same department	1.85	1.98	1.82	1.79	1.75	1.71	1.73	1.81 p=ns
In other departments of same institution	1.82	1.40	2.21	1.62	1.45	1.33	1.78	1.59 p=.000
In same discipline in other institutions	2.00	2.08	1.84	1.85	1.85	1.83	1.82	1.92 p=ns

Scale: 1 = lower, 2 = about the same, 3 = higher.

somewhat less importance in introductory courses ($M = 3.01$), carried the most weight in advanced courses ($M = 3.57$). Averaged across the seven fields, oral reports, written assignments, creativity, class participation, group projects, and subject-specific skills and techniques took on increased importance as students moved from introductory to advanced level courses, while tests and quizzes declined in importance.

Departmental variations in the importance of various factors in grading introductory courses are shown in Table 19 along with the probability of the *F*-test ratio for significant differences across fields. Except for the importance of adhering to due dates ($F = 2.0619$, $df = 524$, $p = .0561$), there were statistically significant differences across disciplines for each of these factors. However, although departments showed significant variations across fields ($F = 2.7331$, $df = 524$, $p = .0000$), the Scheffe procedure showed no pairs of departments that differed significantly with respect to these factors.

Tests and quizzes varied in importance ($F = 11.9438$, $df = 530$, $p = .0000$), with faculty in English placing significantly less importance on tests when determining grades in introductory courses than faculty in business, chemistry, history, mathematics, and psychology; faculty in education placed significantly less importance on tests than faculty in business, chemistry, mathematics, and psychology. There was also significant variation across fields ($F = 27.4984$, $df = 527$, $p = .0000$) in the importance of papers, with faculty in mathematics, chemistry, and business giving significantly less importance to papers than faculty in English, education, and history. Oral reports also showed significant differences across fields ($F = 9.4922$, $df = 503$, $p = .0000$), with education faculty giving them significantly more importance than faculty in business, chemistry, English, history, mathematics, and psychology, and faculty in English giving them significantly more importance than faculty in chemistry. A similar pattern occurred with respect to group projects ($F = 12.1219$,

TABLE 18

Importance Given to Various Factors in Assigning Grades in Introductory and Advanced Courses

	Introductory	*Advanced*	*Difference*
Tests and quizzes	3.61	3.38	-.23
Papers and written assignments	3.01	3.57	.56
Skills and techniques	2.82	3.05	.23
Adherence to due dates	2.69	2.74	.05
Creativity	2.45	2.91	.46
Attitude and effort	2.37	2.52	.15
Improvement	2.35	2.31	-.04
Ability level	2.27	2.37	.10
Departmental norms and standards	2.24	2.15	-.09
Class participation	2.23	2.61	.38
Attendance	2.04	2.11	.07
Personal circumstances that may have affected academic performance	1.92	1.91	-.01
Oral reports	1.62	2.38	.76
Group projects	1.56	1.93	.37
Background characteristics (e.g., socioeconomic status) that may have affected academic performance	1.54	1.51	-.03

Scale: 1 = no importance, 2 = slight importance, 3 = moderate importance, 4 = great importance.

BEST COPY AVAILABLE

TABLE 19

Departmental Means for Importance Given to Various Factors When Assigning Grades in Introductory Courses and Probability of *F*-Test Ratio for Significant Differences Across Fields

	Business	*Chemistry*	*Education*	*English*	*History*	*Mathematics*	*Psychology*	*Total*
Tests and quizzes	3.81	3.87	3.22	2.98	3.61	3.93	3.86	3.61 p=.000
Papers and written assignments	2.72	2.46	3.50	3.76	3.40	2.51	2.76	3.01 p=.000
Skills and techniques	2.83	2.69	2.94	3.09	2.94	3.09	1.90	2.82 p=.000
Adherence to due dates	2.80	2.48	2.66	2.69	2.95	2.69	2.59	2.69 p= ns
Creativity	2.20	2.15	2.81	2.90	2.92	2.13	2.08	2.45 p=.000
Attitude and effort	2.52	2.04	2.69	2.84	2.63	2.00	2.06	2.37 p=.000
Improvement	2.40	2.09	2.59	2.73	2.72	1.97	2.06	2.35 p=.000
Ability level	2.17	2.04	2.00	2.77	2.60	2.10	1.87	2.27 p=.000
Departmental standards	2.09	2.49	2.00	2.33	2.19	2.26	1.97	2.24 p=.013
Class participation	2.51	1.64	2.78	2.83	2.65	1.85	1.67	2.23 p=.000
Attendance	2.15	1.51	2.69	2.64	2.42	1.65	1.55	2.04 p=.000
Personal circumstances	2.00	1.74	2.19	2.17	1.99	1.70	1.86	1.93 p=.000
Oral reports	1.75	1.38	2.45	1.81	1.58	1.42	1.40	1.62 p=.000
Group projects	1.81	1.32	2.61	1.67	1.29	1.38	1.52	1.56 p=.000
Background characteristics	1.35	1.33	1.84	1.86	1.73	1.28	1.52	1.55 p=.000

Scale: 1 = no importance, 2 = slight importance, 3 = moderate importance, 4 = great importance.

df = 490, *p* = .0000), with education faculty giving them significantly more importance than faculty in business, chemistry, English, history, mathematics, and psychology.

Attendance also varied considerably in importance as a component of introductory course grades (F = 22.3079, *df* = 524, *p* = .0000), with chemistry faculty giving it significantly less importance than business, education, English, and history faculty. Mathematics and psychology faculty placed significantly less importance on attendance than did faculty in education, English, and history. Class participation was significantly more important as a component of grades for introductory courses in English, education, history, and business than in chemistry, psychology, and mathematics (overall F = 29.8559, *df* = 526, *p* = .0000). Psychology faculty placed significantly less importance on skills and techniques than faculty in the other six fields; the F-ratio across departments for the importance of skills and techniques was 12.7782 (*df* = 500, *p* = .0000).

Faculty in English and history placed significantly more importance on student improvement when assigning grades in introductory courses than did faculty in mathematics, psychology, and chemistry; student improvement also showed significant variation across disciplines (F = 11.0496, *df* = 526, *p* = .0000). Creativity was significantly more important in English and history than in psychology, mathematics, chemistry, and business, and significantly more important in education than in psychology and mathematics. The importance of creativity varied significantly across fields (F = 14.4839, *df* = 510, *p* = .0000). The importance assigned to students' attitude and effort varied considerably across departments (F = 10.7004, *df* =522, *p* = .0000), with faculty in English, education, and history placing more importance on this factor than faculty in mathematics and chemistry; faculty in English and history also gave significantly more importance to attitude and effort than faculty in psychology.

The ability level of students played an important role in introductory course grades, varying significantly

TABLE 20

Factors with Varying Levels of Importance in Assigning Grades in Introductory Undergraduate Courses

	Very Great Importance (3.50 or higher)	*Great Importance (3.00–3.49)*	*Moderately High Importance (2.75–2.99)*	*Moderate Importance (2.50–2.74)*
Business	Tests	—	Skills Due dates	Papers Attitude Participation
Chemistry	Tests	—	—	Skills
Education	Papers	Tests	Skills Creativity Participation	Attendance Attitude Due dates Group projects Improvement
English	Papers	Skills	Tests Creativity Attitude Participation Ability level	Improvement Due dates
History	Tests	Papers	Due dates Skills Creativity	Improvement Participation Attitude Ability level
Mathematics	Tests	Skills	—	Due dates Papers
Psychology	Tests	—	Papers	Due dates

Scale: 1 = no importance, 2 = slight importance, 3 = moderate importance, 4 = great importance.

across fields ($F = 7.8582$, $df = 495$, $p = .0000$), with faculty in English considering ability more important than those in psychology, chemistry, and mathematics, and faculty in history giving it more importance than faculty in psychology. Faculty also varied in the extent to which they considered student background characteristics ($F = 9.6976$, $df = 523$, $p = .0000$), with those in English giving them significantly more importance than those in mathematics, chemistry, and business, those in education giving them more importance than those in mathematics, and those in history giving them more importance than those in mathematics and chemistry. There was also significant variation in the extent to which faculty considered students' personal circumstances when assigning grades ($F = 5.0892$, $df = 523$, $p = .0000$), with faculty in English significantly more likely to consider them when grading than faculty in mathematics and chemistry.

Table 20 summarizes the factors faculty considered most important in introductory course grades across fields. Chemistry, psychology, and mathematics faculty tended to adopt relatively simple models for grading introductory courses, with tests taking on high importance and only one or two other factors, such as subject-specific skills, also considered. In contrast, education, English, and history faculty adopted rather complex models for grading introductory courses. In both education and English, papers and other written work were considered more important than tests and, in addition, eight or nine other factors were included in grading in these subjects. Creativity, improvement, and class participation were moderately important in grades in education, English, and history. Business tended to fall between these two groups in the factors faculty considered important in grading introductory courses.

Departmental variations in the importance of these same factors in grading advanced-level courses are shown in Table 21 along with the probability of significant differences across fields. Papers and other written assignments ($M = 3.52$) superseded tests and quizzes ($M = 3.38$) as the most important factor in grading advanced courses.

All the factors showed significant variation for advanced courses across the seven fields. However, the importance of adherence to due dates ($F = 2.5462$, $df = 539$, $p = .0194$) showed no significant differences between pairs of fields. Tests and quizzes had significantly more importance for grades in advanced courses in chemistry and mathematics than in English, education, and history. Business and psychology faculty also placed significantly more importance on tests and quizzes than faculty in English. Variation in the importance of tests and quizzes across departments yielded an F of 13.8415, $df = 540$, $p = .0000$. The pattern for the importance of papers and written assignments was almost the reverse, with English, education, history, and psychology faculty considering these more important than mathematics or chemistry faculty; the over-all F

TABLE 21

Departmental Means for Importance Given to Various Factors When Assigning Grades in Advanced Courses and Probability of *F*-Test Ratio for Significant Differences Across Fields

	Business	*Chemistry*	*Education*	*English*	*History*	*Mathematics*	*Psychology*	*Total*
Tests and quizzes	3.40	3.78	3.14	2.91	3.17	3.75	3.44	3.38 *p*=.000
Papers and written assignments	3.53	3.13	3.80	3.84	3.76	3.10	3.66	3.52 *p*=.000
Skills and techniques	3.02	2.91	3.28	3.22	3.23	3.23	2.45	3.05 *p*=.000
Adherence to due dates	3.00	2.57	2.66	2.68	3.00	2.67	2.70	2.75 *p*=.019
Creativity	2.83	2.64	2.94	3.20	3.18	2.75	2.88	2.92 *p*=.000
Attitude and effort	2.64	2.29	2.78	2.88	2.68	2.07	2.45	2.52 *p*=.000
Improvement	2.36	2.12	2.48	2.60	2.61	1.93	2.12	2.31 *p*=.000
Ability level	2.38	2.09	2.08	2.86	2.67	2.24	2.13	2.37 *p*=.000
Departmental standards	1.94	2.45	2.10	2.25	2.12	2.08	1.86	2.14 *p*=.002
Class participation	3.01	2.01	2.78	2.99	3.05	2.07	2.67	2.61 *p*=.000
Attendance	2.23	1.62	2.60	2.61	2.45	1.62	1.85	2.11 *p*=.000
Personal circumstances	1.96	1.70	2.12	2.15	1.97	1.68	1.92	1.92 *p*=.000
Oral reports	2.58	2.32	2.69	2.35	2.37	1.86	2.67	2.38 *p*=.000
Group projects	2.76	1.71	2.67	1.77	1.54	1.64	2.04	1.93 *p*=.000
Background characteristics	1.35	1.28	1.70	1.79	1.73	1.24	1.55	1.52 *p*=.000

Scale: 1 = no importance, 2 = slight importance, 3 = moderate importance, 4 = great importance.

was 16.0054, *df* = 545, p = .0000. Oral reports (*F* = 5.4873, *df* = 527, *p* = .0000) took on significantly more importance for grades in education, psychology, and business than in mathematics. Group projects (*F* = 14.3659, *df* = 502, *p* = .0000) were more important in business and education than in history, mathematics, chemistry, and English; they were also significantly more important in business than in psychology.

Attendance (*F* = 16.7542, *df* = 538, *p* = .0000) was significantly more important as a component of advanced course grades in English, education, history, and business than in chemistry and mathematics; attendance was also significantly more important in English, education, and history than in psychology. Class participation (*F* = 19.7710, *df* = 541, *p* = .0000) was significantly less important in advanced courses in chemistry and mathematics than in business, history, English, and education. Skills and techniques (*F* = 7.5818, *df* = 522, *p* = .0000) were a significantly more important component of grades for advanced courses in education, history, mathematics, and English than in psychology. Improvement (*F* = 6.9727, *df* = 539, *p* = .0000) was more important for grades in history and English than in mathematics, and more important in English than business. Creativity (*F* = 4.7615, *df* = 531, *p* = .0000) was significantly more important for grades in English and history than in chemistry.

Attitude and effort (*F* = 8.3919, *df* = 542, *p* = .0000) was significantly more important in English, education, history, and business than in mathematics, and more important in English than in chemistry. Faculty in English placed significantly more importance on students' ability level when grading advanced courses than did faculty in education, chemistry, psychology, and mathematics; the *F*-ratio was 6.1851 (*df* = 515, *p* = .0000). The importance given to students' background characteristics also varied (*F* = 8.8193, *df* = 536, *p* = .0000), with faculty in English, history, and education considering this factor more important than those in mathematics, faculty in English and history consid-

TABLE 22

Factors with Varying Levels of Importance in Assigning Grades in Advanced Undergraduate Courses

	Very Great Importance (3.50 or higher)	*Great Importance (3.00–3.49)*	*Moderately High Importance (2.75–2.99)*	*Moderate Importance (2.50–2.74)*
Business	Papers	Tests Skills Participation Due Dates	Creativity Group projects	Attitude Oral reports
Chemistry	Tests	Papers	Skills	Creativity
Education	Papers	Skills Tests	Creativity Attitude Participation	Oral reports Group projects Due dates Attendance
English	Papers	Skills Creativity	Participation Tests Attitude Ability level	Due dates Improvement
History	Papers	Skills Creativity Tests Participation Due dates	—	Attitude Improvement
Mathematics	Tests	Skills Papers	Creativity	Due dates
Psychology	Papers	Tests	Creativity	Due dates Oral reports Participation

Scale: 1 = no importance, 2 = slight importance, 3 = moderate importance, 4 = great importance.

ering it more important than those in chemistry, and faculty in English considering it more important than those in business. Personal circumstances of students also had a varying influence on grades in advanced courses ($F = 5.0382$, $df = 537$, $p = .0000$), with English faculty considering this factor significantly more important than faculty in mathematics or chemistry. Finally, departmental standards also had a variable influence on advanced course grades ($F = 3.5680$, $df = 539$, $p = .0000$), with chemistry faculty placing significantly more important on this factor than faculty in psychology.

Table 22 summarizes, across fields, the factors faculty considered most important in grading advanced-level courses. Chemistry, mathematics, and psychology faculty still tended to use less complex models for grading, but even these were more complex than for introductory courses. In chemistry and mathematics, tests remained dominant and class participation was still unimportant. Psychology faculty emphasized papers and gave some attention to class participation. Creativity, which was important only in education, English, and history at the introductory level, now became important in every discipline. Education, English, and history faculty continued to use complex grading models, but they were joined by business. In business and history, papers replaced tests as the most important factor in grading. Oral reports, which were not given much importance in grading introductory courses, took on some importance in grading advanced courses in business, education, and psychology; business joined education in using the results of group projects to make grading decisions.

Because some fields give considerably more importance to laboratory work, internships, or fieldwork than do others, faculty were asked if they taught any courses in which at least half the final grade was based on fieldwork, internship, or laboratory work. Fourteen percent indicated that they taught such courses. There were significant differences across fields ($F = 16.3851$, $df = 577$, $p = .0000$). Forty-three percent of education, 26 percent of psychology, 19 percent of chemistry, and 14 percent of business faculty said they taught such courses. Education faculty taught this type of course significantly more frequently than did faculty in any other field; chemistry and psychology faculty taught this type of course significantly more frequently than did faculty in English, history, or mathematics.

Faculty who reported teaching laboratory, fieldwork, or internship courses were asked to indicate the importance of each of the grading factors in these courses in comparison to their importance in other courses. The results are summarized in Table 23. Caution is necessary when considering these results since the responses are based on approximately 75 faculty, primarily faculty in education, psychology, chemistry, and

TABLE 23

Importance Given to Various Factors in Assigning Grades in Laboratory, Fieldwork, and Internship Courses Versus Other Courses

	Laboratory Fieldwork/Internship	*Other*	*Difference*
Tests and quizzes	2.08	3.44	−1.36
Papers and written assignments	3.41	3.51	−0.10
Skills and techniques	3.42	2.81	0.61
Adherence to due dates	3.17	2.94	0.23
Creativity	2.88	2.70	0.18
Attitude and effort	3.18	2.44	0.74
Improvement	2.77	2.40	0.37
Ability level	2.22	2.00	0.22
Departmental norms and standards	2.23	2.26	−0.03
Class participation	2.93	2.52	0.41
Attendance	3.21	2.27	0.94
Personal circumstances that may have affected academic performance	1.95	1.96	−0.01
Oral reports	2.49	2.46	0.03
Group projects	2.22	2.23	−0.01
Background characteristics (e.g., socioeconomic status) that may have affected academic performance	1.53	1.49	0.04
Requirements for profession	2.75	2.27	0.46

Scale: 1 = no importance, 2 = slight importance, 3 = moderate importance, 4 = great importance.
N averages about 75.

business. Two factors had the greatest importance in grading laboratory and fieldwork courses—skills and techniques and papers and written assignments. Attendance, attitude and effort, and adherence to due dates were also given great importance. Class participation, creativity, improvement, and requirements for the profession (a grading factor not included for introductory or advanced courses) appeared to have moderately high importance in such courses. The greatest difference between laboratory and fieldwork courses and other courses was the lower importance given to tests and quizzes. In these courses, faculty gave greater importance to attendance, attitude and effort, skills and techniques, requirements for the profession, class participation, and improvement than they did in their other courses. There was significant variation across departments regarding the requirements of the profession when grading laboratory/fieldwork courses ($F = 3.0421$, $df = 74$, $p = .0154$), with education faculty placing significantly more importance on this factor than faculty in other fields.

Assessment in Introductory Courses

Faculty were asked about the type(s) of assessment they used in their introductory courses, who prepared the tests, and the source(s) for test content.

As Table 24 shows, responding faculty were more likely to use essay examinations than multiple-choice tests. The types of assessment used differed significantly across fields. Essay tests ($F = 32.9212$, $df = 445$, $p = .0000$) were used significantly more frequently by English and history faculty than chemistry, mathematics, and business faculty; they were used significantly more frequently by education and psychology faculty than by chemistry and mathematics faculty, and significantly more frequently by business faculty than by mathematics faculty. Multiple-choice tests ($F = 39.8679$, $df = 466$, $p = .0000$) were used significantly more frequently by psychology and business faculty than by English and mathematics faculty; psychology faculty used multiple-choice tests significantly more frequently than chemistry and history faculty. Short-answer tests ($F = 2.9032$, $df = 454$, $p = .0087$) were used significantly more frequently by chemistry faculty than English faculty. Performance assessment ($F = 6.6893$, $df = 372$, $p = .0000$) was used significantly more frequently by education, mathematics, and business faculty than by history faculty.

When asked who prepared the tests for introductory courses, about 77 percent of faculty stated they prepared their tests alone. About 8 percent prepared tests for introductory courses with other faculty who also taught these courses. About 6 percent of faculty who taught introductory courses involved a teaching assistant in the preparation of examinations.

About 30 percent of responding faculty said they got the items for the tests in their introductory courses from a textbook publisher; about 8 percent said they got their items from a departmental item bank. There

TABLE 24

Types of Assessment Used in Introductory Undergraduate Courses and Probability of *F*-Test Ratio for Significant Differences Across Fields

	Percentage of Faculty Using Each Type				
	Multiple Choice	*Short Answer*	*Essay*	*Performance*	*Other*
Business	43.3	23.7	24.5	44.0	27.1
Chemistry	35.8	39.0	21.5	26.0	48.6
Education	38.2	31.3	32.1	26.9	26.8
English	16.2	24.4	69.7	15.5	22.2
History	26.2	28.6	67.3	13.9	22.8
Mathematics	17.5	38.6	31.3	65.7	74.1
Psychology	57.3	33.6	29.4	20.5	22.4
Total	37.0 *p*=.0000	31.9 *p*=.0087	47.4 *p*=.0000	35.0 *p*=.0000	52.5 *p*=.0000

were significant differences across fields in the use of these item sources. Items from textbook publishers (F = 27.6312, df = 562, p = .0000) were used significantly more frequently by psychology faculty than by faculty in English, history, mathematics, and chemistry; business faculty were significantly more likely to use publishers' items than faculty in English, history, mathematics, and chemistry, while chemistry and education faculty were significantly more likely to use such items than faculty in English. Departmental item banks (F = 5.2571, df = 522, p = .0000) were used significantly more frequently by chemistry faculty than by faculty in English or history.

Student Evaluations and Challenges to Grading

A final group of questions dealt with pressures that may make faculty consider modifying or changing their grading standards. These pressures may come from the use of student course evaluations, either by the department or by the institution, and/or from challenges to the grades faculty have given.

Nearly all (99 percent) of responding faculty indicated that student evaluations of faculty took place in their courses. Faculty reported that these evaluations were used for tenure decisions (83 percent), promotion decisions (81 percent), retention decisions (76 percent), and merit pay decisions (57 percent), as well as to help improve instruction (77 percent). While most faculty said they seldom or never considered their reputation with students and how they would be evaluated by them when assigning course grades (34 percent said they did not consider these factors, 35 percent said they gave them very little consideration), other faculty were more concerned. Twenty-six percent of faculty said they sometimes gave these factors some consideration, while 6 percent said they gave them a lot of consideration.

Sometimes faculty are challenged about or pressured to change a grade. In this study, about 65 percent of responding faculty said students seldom or never challenged a grade, while about 30 percent said they had such challenges from one or two students in each class they taught and about 3 percent said they were usually challenged by about 5 percent of the students in a class. Only about 12 percent of the faculty said that, when challenged, they usually changed the student's grade. There was significant variation across departments in the extent of student challenges (F = 5.4119, df = 568, p = .0000), with chemistry faculty being challenged significantly more frequently than faculty in English or mathematics. Relatively few (13 percent) of faculty said they had ever been asked by a member of the faculty or administration to change a student's grade.

Discussion and Conclusions

This report summarizes institutional policies related to grading at 14 colleges and universities and how these policies changed between 1980 and 1990. The report also summarizes grading policies and practices in the business, chemistry, education, English, history, mathematics, and psychology departments at these institutions and the grading orientation and practices of faculty in these departments.

Changes Between 1980 and 1990

Information about grading policies and practices was obtained from institutional, department chair, and faculty questionnaires. Four areas of change that may be

related to grading were identified: (1) changes in curriculum requirements; (2) changes in grading systems; (3) increased use of student evaluations to make decisions about faculty; and (4) changes in faculty grading philosophies and attitudes.

The institutional questionnaires indicated that current curriculum requirements at these institutions tended to be more prescribed than they had been a decade ago. Seven of the 14 institutions had added general education courses that all students were required to take. This finding, which is similar to those of other surveys (e.g., Lewis and Farris 1989; Toombs and Fairweather 1989), suggests that changes in general education requirements during the 1980s may have affected the grades students received, especially in the first two years of college when general education courses are most frequently taken. Students may be less motivated to achieve in required courses than in elective courses, which they may find more interesting. Also, the required curriculum limits students' opportunities to take "easy" courses. Institutional practices may also affect the grades in required general education courses. For example, required courses may have larger class sizes than other courses and/or may be taught by less experienced faculty. As noted in the review of relevant literature, both class size and faculty teaching experience have been found to be inversely related to strict grading practices. The reported curriculum changes may be contributing to lower GPAs at the institutions in this survey.

The institutional questionnaires also indicated that grading systems at these institutions tended to be more differentiated in 1990 than they had been in 1980. Eight of the 14 colleges and universities were using a letter-grade system with pluses and minuses in 1990, a system used by only 4 of these institutions in 1980. At one university the change to a more differentiated system had been mandated by the administration as part of an effort to combat grade inflation. This change appears to have produced the desired results at this institution but it is unclear whether it has had similar effects at the other colleges and universities studied. In earlier research on this topic, Potter (1979) found that adding pluses and minuses to letter grades resulted in higher GPAs than were produced by the simpler five-letter grading system.

Still another area of change that may have affected grades is the use of student evaluations of faculty to make decisions about faculty retention, promotion, and tenure (and, less often, salaries). Eleven of the 14 colleges and universities indicated that such evaluations were required institution-wide in 1990, while only 8 had used such evaluations in 1980. In each case the institution indicated that these evaluations were used in retention, promotion, and tenure decisions. Faculty questionnaires also indicated wide use of student evaluations; 99 percent of responding faculty said they were used in their department and more than 75 percent said these evaluations affected tenure, promotion, and retention decisions as well as helping the department improve instruction. This extensive use of student evaluations to make major decisions about faculty, especially during a period when faculty positions have become increasingly difficult to find and keep, raised the question that some faculty might be afraid to grade too harshly lest they receive unfavorable student reviews and jeopardize their careers. This concern was confirmed in the faculty questionnaire. Nearly a third (32 percent) of responding faculty said that, when assigning grades, they gave a lot or some consideration to how they would be evaluated by their students. This policy may contribute unwittingly to grade inflation.

Faculty members were asked directly to indicate what they thought were the three most important reasons for any grade inflation that occurred between 1970 and 1990. Thirty percent of the respondents said that the most important reason was that faculty now expect less of students; 54 percent said this was one of the three most important reasons. Other frequently reported reasons were that faculty now have less concern for maintaining high standards and that faculty have changed their grading focus away from absolute standards and toward student improvement. An estimate of the extent to which changes in faculty grading philosophy may have occurred can be determined by the fact that slightly more than 40 percent of responding faculty indicated they had changed their philosophy since they began teaching.

Institutional and Departmental Influences on Grades

This study found that relatively few of the colleges and universities surveyed had prescribed grading policies. Of those that did, the most common policy was grading against specific standards. This policy was reported at three institutions and had not changed between 1980 and 1990. According to the department chairs, there were no departmental policies regarding grading practices. However, faculty had a different perspective. Approximately a quarter (26 percent) of responding faculty said their department had a policy requiring grading against specific standards. One possible reason for this discrepancy is that faculty believed that institutional policies requiring grading against standards (in place at three institutions) were departmental policies.

Approximately 11 percent of faculty reported a departmental policy concerning including attendance in the course grade, a figure that also suggests confusion with the institutional policy concerning this factor at one college. However, while no institution or department had a policy about grades in honors courses, 14 percent of faculty believed that there was such a departmental policy.

Faculty were much more likely to report that certain grading practices were expected in their department. Nearly two-thirds (63 percent) of responding faculty said their department expected them to grade against specific standards. About half (51 percent) of faculty said their department expected that they would grade students relative to the overall performance of the class and nearly half (47 percent) said their department expected them to give a lower grade if students were late in turning in work. Between a quarter and a third of responding faculty indicated their departments expected that students in honors courses would get an A or a B, that attendance and/or effort would be included in the course grade, that individual progress would be included in the course grade, that students would be graded "on the curve," and/or that attitude and behavior would be included in the course grade. Departmental chairs probably communicated such expectations, either in discussions about grading or in conjunction with their review of course grades. Departmental review of grades was reported in all the less selective institutions; such review was least common in the more selective institutions. In addition, all the department chairs in less selective institutions reported holding informal discussions about grading. These findings suggest that departments in the less selective institutions are the most likely to be concerned about maintaining expected grading practices.

Overview of Faculty Grading Philosophies and Practices

Overall, responding faculty were most likely to say that they emphasized formal, objective measurement of knowledge and that they graded students in terms of how their achievement level compared with some absolute standard. Faculty gave the most importance to tests and quizzes when they assigned grades in their introductory courses and were most likely to use essay tests when assessing students in these courses. In advanced courses, faculty tended to give more importance to papers and written assignments than to tests and quizzes. Those faculty who taught laboratory or fieldwork courses also tended to place more importance on papers and written assignments than on tests and quizzes. Faculty members responding to this survey believed that the most important purpose of grading was to give feedback to students.

Departmental Variations in Grading

This study confirmed what other studies have found—there are many significant differences across departments in grading philosophies and practices. Faculty differed significantly in their attitudes about grading, with those in the sciences more likely to say they emphasized formal objective measurement of knowledge than those in the humanities or in preprofessional areas (such as business and education). However, there were no significant differences across departments in the percentage of faculty who said they use or preferred a criterion-referenced approach to grading.

There were significant differences in the importance that faculty in the different departments assigned to various grading factors, both for introductory and for advanced courses. In introductory courses, faculty in business, chemistry, history, mathematics, and psychology considered tests and quizzes most important, while faculty in education and English considered papers and written assignments most important. Explanations for these preferences may be associated with the nature of the subject or with other characteristics of the course, such as typical class size. Such preferences appear to be linked to the type of assessment used. In introductory courses, psychology faculty were most likely to use multiple-choice tests, while English and history faculty tended to use essays and mathematics faculty used performance tests. In advanced courses, chemistry and mathematics faculty gave the most importance to tests and quizzes, while faculty in the other departments considered papers and written assignments most important.

The importance of tests for grades in introductory chemistry, history, and psychology courses may be related, in part, to the relatively large size of introductory courses in these subjects. Similarly, the relatively small size of introductory English classes, as well as the nature of the subject, may explain the greater emphasis given to papers and written assignments. The smaller size of advanced-level courses may also account for the shift in emphasis from tests to papers and written assignments by faculty in business, history, and psychology. Both chemistry and mathematics, which use relatively simple grading models that emphasize tests at both the introductory and advanced course levels, are subjects in which reduction of information by quantification plays

an important role. In contrast, education and English, which use more complex grading models that emphasize papers and written assignments at both levels, are subjects dealing with behaviors that are considered to have complex explanations that cannot be readily quantified.

There were fewer significant differences in faculty opinions about the purposes of grading than in grading philosophies and practices; faculty in mathematics and history were more likely to view grades as a way of motivating students than were faculty in education and psychology.

Additional Analyses and Future Research

There are several additional analyses that could be carried out with this data to provide further information about how the grading process varies and about how it may have changed.

It would be highly desirable to analyze this data by institutional characteristics. As indicated in the literature review, grades tend to be slightly less predictable in institutions that enroll students with a wide range of academic ability, once there is a correction for restriction of range; grades also may vary more in institutions with a very diverse curriculum. In addition, variations in faculty grading philosophies and practices in more and less selective institutions need to be explored.

It would also be useful to compare the current grading philosophies and practices of faculty who say they have and have not changed their approach to grading since they began teaching. Such an analysis would help confirm the hypothesis that faculty are now more focused on student improvement and less on absolute standards than they were in the past; it would also help confirm the argument that faculty are now less concerned with high standards. Finally, such an analysis might shed some light on the changing expectations that faculty have for students.

The multi-level nature of these data also needs to be explored further. For example, to what extent do faculty grading practices differ when the use of specific grading standards is mandated by the institution or is an expected practice of the department? How do faculty grading approaches vary in departments that do and do not review the grades their faculty give; are such variations related to the use of part-time faculty? Also, it would be interesting to do some correlational analyses of these data to determine how the different components are interrelated. This might provide the groundwork for a future study to determine the extent to which grades are affected by institutional, departmental, faculty, and student characteristics.

Future studies of grading should be designed so that the findings can be linked to actual grade records. It would be especially useful if comparisons could be made between institutions that have shown higher and lower rates of grade inflation.

Grading in the Mid-1990s

Does this report provide any clues about whether and how the predictability of the GPA will change? The answer is "yes, there are clues," but the clues are somewhat contradictory.

There appears to be considerable pressure on institutions of higher education and their faculties to reduce what the public perceives as lax standards resulting in ever rising GPAs. Institutions seem to be taking steps to respond to this, primarily by introducing curriculum requirements. At the departmental level, slightly more than half the responding faculty reported efforts to raise standards. There also appears to be informal pressure for faculty to meet certain expected departmental grading standards, especially at the less selective institutions in this survey. Both of these trends might be expected to lead to lower overall GPAs and to make the GPA more predictable.

However, other forces are at work that may have the opposite effect. Many institutions are using a more differentiated grading system. This was shown, in the past, to be associated with inflation of the GPA. It is not entirely clear from these data whether or not greater differentiation is having this effect at the institutions in this survey that have adopted a system of letter grades with pluses and minuses. Another, potentially more powerful force for grade inflation is the increased use of student course evaluations to make decisions about faculty careers. As we move through the 1990s, competition to enroll and retain students appears to be a matter of increasing concern in higher education. Faculty at institutions most in danger of loosing students who receive low grades are likely to feel increased pressure to modify grades sufficiently so that students will reenroll. These faculty members are also likely to be concerned about their own futures and, especially among junior faculty, feel it necessary to adapt their grading standards to their students. Nearly a third of responding faculty said they gave some or a lot of consideration to the impact on student evaluations when assigning course grades. Nearly half of responding faculty said they had changed their grading philosophy since they began to teach. Finally, about half said that grade infla-

tion had occurred because faculty today expect less of students than they did in the past.

Just how these competing forces will affect college grades in the middle and late 1990s remains to be seen.

References

Aiken, L.R. 1963. "The Grading Behavior of a College Faculty." *Educational and Psychological Measurement* 23:319–22.

Aleamoni, L.M. 1987. "Student Rating Myths Versus Research Facts." *Journal of Personnel Evaluation in Education* 1:111–19.

Baird, L.L. 1983. *Predicting Predictability: The Influence of Student and Institutional Characteristics on the Prediction of Grades.* College Board Report No. 83-5. New York: College Entrance Examination Board.

Bejar, I.I., and E.O. Blew. 1981. "Grade Inflation and the Validity of the Scholastic Aptitude Test." *American Educational Research Journal* 10:143–56.

Bellott, F. K. 1981. "Relationships of Declining Test Scores and Grade Inflation." Paper presented at the annual meeting of the Mid-South Educational Research Association, November (ED 304 449), Lexington, Ky.

Birnbaum, R. 1977. "Factors Related to Grade Inflation." *Journal of Higher Education* 48:519–39.

Blackburn, R., E. Armstrong, C. Conrad, J. Didham, and T. McKune. 1976. *Changing Practices in Undergraduate Education: A Report for the Carnegie Council on Policy Studies in Higher Education.* Berkeley, Calif.: Carnegie Foundation for the Advancement of Teaching.

Bogart, Q.J., and K.M. Kistler. 1987. "California Community College and California State University English Faculty Grading Practices: An Assessment." *Community/Junior College Quarterly* 11:39–45.

Boli, J., H. Katchadourian, and S. Mahoney. 1984. *Grades, Grading Standards, and Academic Awards.* Stanford, Calif.: Stanford University, Office of Undergraduate Research (ED 248 790).

Boozer, B. 1977. *Grades and Teacher Characteristics.* Albany: State University of New York, Research Report No. 143 (ED 163 064).

Bowles, S., and H. Gintis. 1976. *Schooling in Capitalist America.* New York: Basic Books.

Bromley, D.G., B.C. Busching, D.L. Oliver, and M.S. Szozda. 1981. "Evaluation Practices: A Study of College Grading Using Normative and Utilitarian Orientations." *Teaching Sociology* 8:423–41.

Carnegie Foundation for the Advancement of Teaching. 1989. *The Condition of the Professoriate: Attitudes and Trends.* Princeton, N.J.: Carnegie Foundation for the Advancement of Teaching.

Carney, P., R.L. Isakson, and R. Ellsworth. 1978. "An Exploration of Grade Inflation and Some Related Factors in Higher Education." *College and University* 53:217–30.

Chase, C. I., and L.M. Wakefield. 1984. *Testing and Grading: Faculty Attitudes and Opinions.* Bloomington, Ind.: Indiana Studies in Higher Education Monograph No. 51, Indiana University (ED 256 196).

Collins, J., and K.N. Nickel. 1979. *A New Look at Grading, Recording, and Averaging Practices in Higher Education in the United States* (ED 176 675).

Dickson, V.A. 1984. "An Economic Model of Faculty Grading Practices." *Journal of Economic Education* 15:197–203.

Dressel, P.L. 1961. "Testing and Grading Policies." In *Evaluation in Higher Education*, ed. P.L. Dressel, pp. 215–32. Boston, Mass.: Houghton Mifflin.

Dressel, P.L. 1976. *Evaluation in Higher Education.* San Francisco: Jossey-Bass.

Dressel, P.L., and F.H. DiLisle. 1969. *Undergraduate Curriculum Trends.* Washington, D.C.: American Council on Education.

Duke, J.D. 1983. "Disparities in Grading Practice, Some Resulting in Inequities, and a Proposed New Index of Academic Achievement." *Psychological Reports* 53:1023–80.

Elliott, R., and A.C. Strenta. 1988. "Effects of Improving the Reliability of the GPA on Prediction Generally and on Comparative Predictions for Gender and Race Particularly." *Journal of Educational Measurement* 25:333–47.

Etzioni, A. 1961. *The Comparative Analysis of Complex Organizations.* New York: Macmillan.

Etzioni, A. 1975. "Grade Inflation." *Science* 190:101.

Felder, N.L. 1979. *The Introduction of a New Grading System at a Public Teaching Institution: Impact on Grading Tendencies of the Faculty.* Ashville, N.C.: University of North Carolina at Ashville (ED 189 997).

Fishman, J.A. 1958. "Unsolved Criterion Problems in the Selection of College Students." *Harvard Educational Review* 28:340–49.

Gallini, J. 1982. "Evidence of an Adaptive Level Grading Practice Through a Causal Approach." *Journal of Experimental Education* 111:188–94.

Geisinger, K.F. 1982. "Marking Systems." In *Encyclopedia of Educational Research*, ed. H.E. Mitzel, 3, 1139–49. New York: Free Press.

Geisinger, K.F., and W. Rabinowitz. 1979. "Grading Attitudes and Practices among College Faculty Members." In *A Spotlight on Educational Problems*, eds., H. Dahl, A. Lyshe, and P. Rand. Oslo, Norway: Universitesforlaaget, pp. 92–105.

Geisinger, K.F., A.N. Wilson, and J.J. Naumann. 1980. "A Construct Validation of Faculty Orientations Toward Grading: Comparative Data from Three Institutions." *Educational and Psychological Measurement* 40:413–17.

Goldman, R.D., D.E. Schmidt, B.N. Hewitt, and R. Fisher. 1974. "Grading Practices in Different Major Fields." *American Educational Research Journal* 11:343–57.

Goldman, R.D., and R.E. Slaughter. 1976. "Why College Grade-Point Average Is Difficult to Predict." *Journal of Educational Psychology* 68:9–14.

Hambleton, R.K., and J. Murray. 1977. "A Comparative Study of Faculty and Student Attitudes Toward a Variety of College Grading Purposes and Practices." Paper presented at the annual meeting of the National Council on Measurement in Education (ED 141 381), April, New York.

Hanna, G.S., and W.E. Cashin. 1988. "Improving College Grading." IDEA Paper No. 19 (ED 298 814). Manhattan, Kan.: Kansas State University.

Haughey, E.J. 1977. "An Analysis of Student and Faculty Attitudes Toward Grading Practices in Relation to Characteristics of Entering Students and Attrition" (ED 150 922). Fort Lauderdale, Fla.: Applied Research Project, Nova University.

Hills, J.R. 1964. "The Effect of Admissions Policy on College Grading Standards." *Journal of Educational Measurement* 2:115–18.

Hills, J.R. 1967. "Housing and Grading Practices." *Florida Journal of Educational Research* 9:57–60.

Hills, J.R., and M.B. Gladney. 1968. "Factors Influencing College Grading Standards." *Journal of Educational Measurement* 5:31–9.

Hills, J.R., and J.A. Klock. 1965. "Predicting Grades for All Schools in a Large University." *American Educational Research Journal* 2:145–50.

Hogrebe, M.C., L. Ervin, P.L. Dwinell, and I. Newman. 1983. "The Moderating Effects of Gender and Race in Predicting the Academic Performance of College Developmental Students." *Educational and Psychological Measurement* 43:523–30.

Hoyt, D.P. 1966. "College Grades and Adult Accomplishments: A Review of Research." *Educational Record*, 47:70–5.

Hoyt, D.P. 1970. "Rationality and the Grading Process." *Educational Record* 51:305–9.

Hughey, J.D., and B. Harper. 1983. "What's in a Grade?" Paper presented at the annual meeting of the Speech Communication Association (ED 248 549), Washington, D.C., November.

Humphreys, L.B. 1968. "The Fleeting Nature of the Prediction of College Academic Success." *Journal of Educational Psychology* 59:375–80.

Johnson, B.G., and H.P. Beck. 1988. "Strict and Lenient Grading Scales: How Do They Affect the Performance of College Students with High and Low SAT Scores?" *Teaching of Psychology* 15:127–31.

Juola, A.E. 1968. "Illustrative Problems in College-Level Grading." *Personnel and Guidance Journal* 29:29–33.

Juola, A.E. 1974. *Grade Inflation (1960–1973). A Preliminary Report.* Ann Arbor, Mich.: Office of Educational Services, Michigan State University.

Kodras, J.E., and J.E. Prather. 1978. *Faculty Opinion on Grading Standards and Practices.* Atlanta, Ga.: Georgia State University, Report No. 78-22, Office of Institutional Planning (ED 160 014).

Lewis, W.A., H.G. Dexter, and W.C. Smith. 1978. "Grading Producers and Test Validation: A Proposed New Approach." *Journal of Educational Measurement* 15:219–27.

Lewis, L.L., and E. Farris. 1989. *Undergraduate General Education and Humanities Requirements.* Higher Education Surveys Report, Survey Number 7, January.

Longstreth, L.E., and D. Jones. 1976. "Some Longitudinal Data on Grading Practices at One University." *Teaching of Psychology* 3:78–81.

Longstreth, L.E. 1979. "Pressure to Reduce Academic Standards." Paper presented at the Annual Meeting of the American Psychological Association, New York.

Mangan, K.S. 1991. "Many Colleges Fill Vacancies with Part-Time Professors, Citing Economy and Uncertainty about Enrollment." *Chronicle of Higher Education* 37, August.

McCornack, R.L., and M.M. McLeod. 1988. "Gender Bias in the Prediction of College Course Performance." *Journal of Educational Measurement* 25:321–31.

McKeachie, W.J. 1986. *Teaching Tips: A Guide for the Beginning College Teacher.* Lexington, Mass.: Heath.

McKenzie, R.B., and G. Tullock. 1981. *The New World Economics.* Homewood, Ill.: Richard D. Irwin.

Miller, R.H. 1965. "Students Show a Preparation Increase but No Increase in Grades Was Shown." *College and University* 41:28–30.

Milton, O., H.R. Pollio, and J. Eison. 1986. *Making Sense of College Grades.* San Francisco, Calif.: Jossey-Bass.

Morgan, R. 1989. *An Examination of the Relationship of Academic Course Work with Admissions Test Performance.* College Board Report 89-6 and ETS Research Report 89-38. New York: College Entrance Examination Board.

Nimmer, J.G., and E.G. Stone. 1991. "Effects of Grading Practices and Time of Rating on Student Ratings of Faculty Performance and Student Learning." *Research in Higher Education* 32:195–15.

Oakes, J. 1985. *Keeping Track: How Schools Structure Inequalities.* New Haven, Conn.: Yale University Press.

Oh, E.K. 1976. "A Study of Instructor Grading Philosophies and Practices at Western Michigan University," doctoral dissertation, Western Michigan University.

O'Leary, B.S. 1980. *College Grade-Point Average as an Indicator of Occupational Success* (PRR-80-23). Washington, D.C.: U.S. Office of Personnel Management.

Ory, J.C., and K.E. Ryan. 1993. *Tips for Improving Testing and Grading.* Newbury Park, Calif.: Sage.

Potter, W.P. 1979. "Grade Inflation: Unmasking the Scourge of the Seventies." *College and University* 55:19–26.

Prather, J.E. 1976. *Faculty Grading Patterns* (Report 59-15). Atlanta, Ga.: Georgia State University, Office of Institutional Research (ED 129 172).

Prather, J.E., G. Smith, and J.E. Kodras. 1979. "A Longitudinal Study of Grades in 144 Undergraduate Courses." *Research in Higher Education* 10:11–24.

Ramist, L., C. Lewis, and L. McCamley. 1990. "Implications of Using Freshman GPA as the Criterion for the Predictive Validity of the SAT." In Willingham, W.W., C. Lewis, R. Morgan, and L. Ramist, *Predicting College Grades: An Analysis of Institutional Trends Over Two Decades.* Princeton, N.J.: College Entrance Examination Board and Educational Testing Service.

Ramist, L., and G. Weiss. 1990. "The Predictive Validity of the SAT, 1964 to 1988." In Willingham, W. W., C. Lewis, R. Morgan, and L. Ramist, *Predicting College Grades: An Analysis of Institutional Trends Over Two Decades.* Princeton, N.J.: College Entrance Examination Board and Educational Testing Service.

Rasbow, J., and A. Hernandez. 1988. "The Price of the 'GPA Perspective': An Empirical Study of 'Making the Grade.'" *Youth and Society* 19:363–67.

Raths, J., M. Wojtaszek-Healy, and C.K. Della-Piana 1987. "Grading Problems: A Matter of Communication." *Journal of Educational Research* 80:133–37.

Rist, R.C. 1970. "Student Social Class and Teacher Expectations: The Self-Fulfulling Prophecy in Ghetto Education." *Harvard Education Review* 40:411–51.

Sabot, R., and J. Wakeman-Linn. 1991. "Grade Inflation and Course Choice." *Journal of Economic Perspectives* 5:159–70.

Sarafina, E.P., and P.A. DiMattia. 1978. "Does Grading Undermine Intrinsic Interest in a College Course?" *Journal of Educational Psychology* 70:916–21.

Stein, S., and W. Guthrie. 1990. "An Analysis of Undergraduate Grades." Research Report. Lawrenceville, N.J.: Rider College.

Strenta, A.C., and R. Elliott. 1987. "Differential Grading Standards Revisited." *Journal of Educational Measurement* 24:281–91.

Suddarth, B.M. 1975. "An Investigation of General Education Requirements in College Curricula." *Research in Higher Education* 3:197–204.

Summerville, R.M., D.R. Ridley, and T.L. Maris. 1988. "Grade Inflation in the Eighties. The Case of Urban Colleges and Universities." Paper presented at the Forum of the Association for Institutional Research, Phoenix, Ariz. (ED 298 869).

Suslow, S. 1977. "Grade Inflation: End of a Trend?" *Change* 9:44–5.

Thorndike, E.L. 1969. "Marks and Marking Systems." In *Encyclopedia of Educational Research*, ed. R. Ebel, pp. 765–66. New York: Macmillan.

Toombs, W., and J. Fairweather. 1989. "Contemporary Profiles of General Education in Four-Year Institutions." Paper presented at the American Educational Research Association annual conference, San Francisco, Calif.

Warren, J.R. 1971. *College Grading Practices: An Overview* RB-71-12. Princeton, N.J.: Educational Testing Service.

Warren, J.R. 1975. *The Continuing Controversy Over Grades* (TM Report 51). ERIC Clearinghouse on Tests, Measurement, and Evaluation. Princeton, N.J.: Educational Testing Service.

Webb, S.C. 1959. "Measured Changes in College Grading Standards." *College Board Review* 39:27–30.

Williamson, L.K., and J. Pier. 1985. "Relationship Between Instructor Degree Status and Student Grades in Basic Speech: A Research Report." Paper presented at the annual meeting of the Speech Communication Association, Denver, Colo. (ED 264 632).

Willingham, W.W. 1985. *Success in College: The Role of Personal Qualities and Academic Abilities*. New York: College Entrance Examination Board.

Willingham, W.W. 1990a. "Introduction: Interpreting Predictive Validity." In Willingham, W.W., C. Lewis, R. Morgan, and L. Ramist. *Predicting College Grades: An Analysis of Institutional Trends Over Two Decades*. Princeton, N.J.: College Entrance Examination Board and Educational Testing Service.

Willingham, W.W. 1990b. "Understanding Yearly Trends: A Synthesis of Research Findings." In Willingham, W.W., C. Lewis, R. Morgan, and L. Ramist. *Predicting College Grades: An Analysis of Institutional Trends Over Two Decades*. Princeton, N.J.: College Entrance Examination Board and Educational Testing Service.

Wilson, K.M. 1970. "Increased Selectivity and Institutional Grading Standards." *College and University* 46:46–53.

Young, J.W. 1993. "Grade Adjustment Methods." *Review of Educational Research* 63:151–65.

Zangenehzadeh, H. 1988. "Grade Inflation: A Way Out." *Journal of Economic Education* 19:217–30.